

T-01

Q統計量における0値の母固有値に 当たる主成分項の従う近似分布



小林靖之・帝京大学 理工学部

- マハラノビス距離では母固有値 $\lambda = 0$ に当たる項を扱えないので、対策の1方法としてQ統計量を用いている。(他には正則化)
- $\lambda = 0$ の判断に理論的基準は無く、標本固有値による経験的判断しかなかった。
 $\lambda = 0$ に当たるQ統計量の主成分項のモデルも無く、マハラノビス距離とQ統計量を融合して同時に扱えなかった。
- 計算機で生じる打切り誤差の影響を考慮すると、
 $\lambda = 0$ を判断する標本固有値の近似モデル(理論的基準の候補)
 $\lambda = 0$ に当たるQ統計量の主成分項の近似モデル
が提案できた。数値実験ではモデルと概ね合う結果を得た。
- 本提案モデルにより、今まで無かったマハラノビス距離とQ統計量の両者を融合したモデルを実現できる可能性があると考えている。

[T-02] Adaptive scaling for soft-thresholding estimator

三重大大学教育学部 萩原克幸

LASSO : ℓ_1 正則化 : **スパース学習**

→ 汎化性とスパース性が必ずしも両立しない

→ SCAD, adaptive-LASSO ... ℓ_1 正則化項の改良

目的 : LASSO 推定量のスケールリングによる改良

直交回帰の場合 : LASSO 推定量 = soft-thresholding 推定量

本発表 : ノンパラメトリック直交回帰の soft-thresholding 推定量に, component-wise かつ data dependent なスケールリングを導入し,

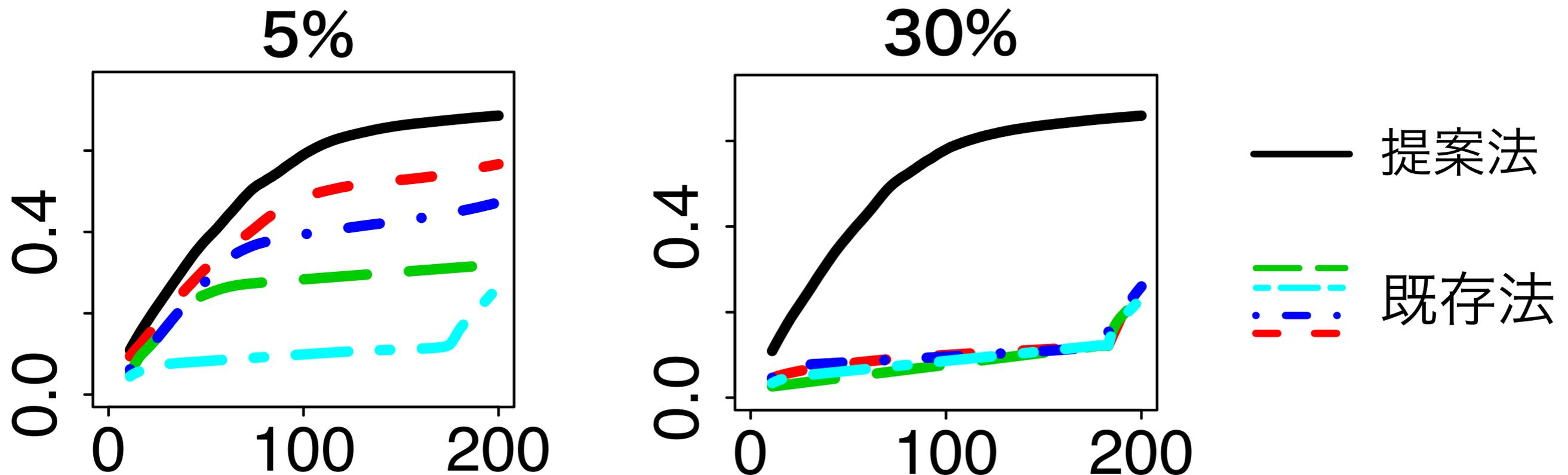
- リスクの不偏推定量を構成 → モデル選択規準
- よりスパースで汎化性の高いモデルを推定できることを数値的に確認 → Wavelet denosing

T-03

スパースガウシアングラフィカル モデルのロバスト推定

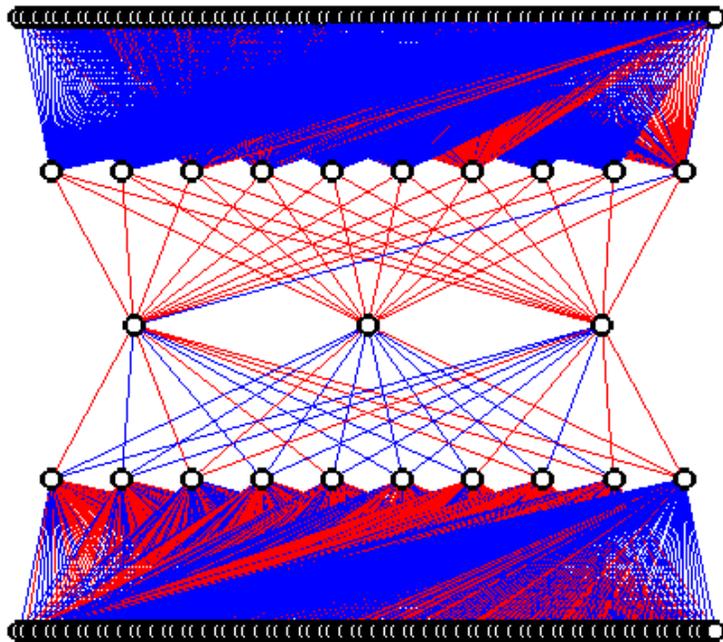
廣瀬慧（大阪大） ・ 藤澤洋徳（統計数理研究所）

- L_1 正則化法によるスパース推定
- γ 尤度法に基づくロバスト推定法を提案
- MMアルゴリズムを導出

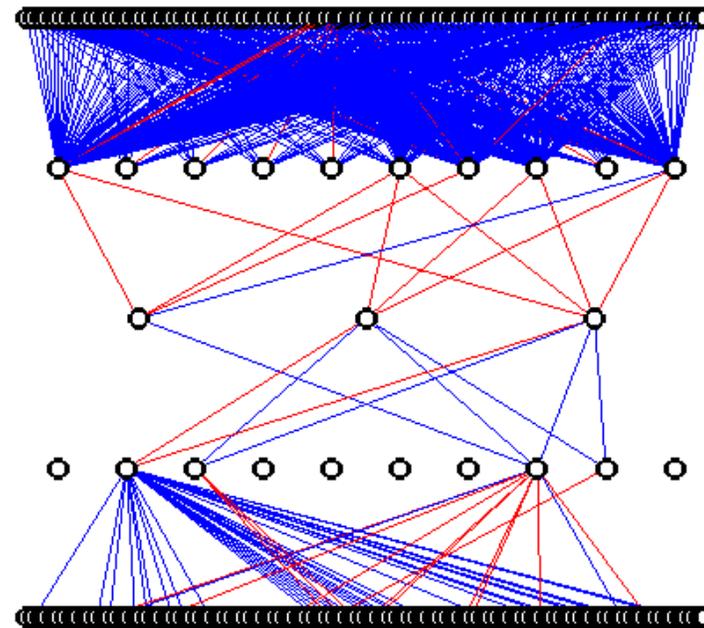


スパース表現を用いた非線形主成分分析 における学習結果の分類法について

T-04 香田 夏輝・渡辺 澄夫(東工大)



従来法の非線形主成分分析



LASSO回帰適用

このスパース表現を
ヒトが理解できる方法
を提案

Wildcardを許容した頻出部分グラフマイニング

T-05

岡崎 文哉 ・ 瀧川 一学 (北大)

背景

- ◆ 頻出部分グラフマイニングは、グラフデータベース中に高頻度で出現する部分グラフを列挙する問題
- ◆ 頻出部分グラフには、ラベルが一部異なるだけの類似部分グラフが多数存在する

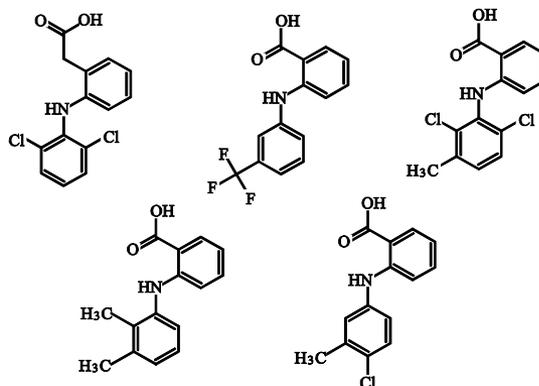
Wildcard許容

- ◆ よりグラフの特徴を表す構造を持つグラフパターンを抽出

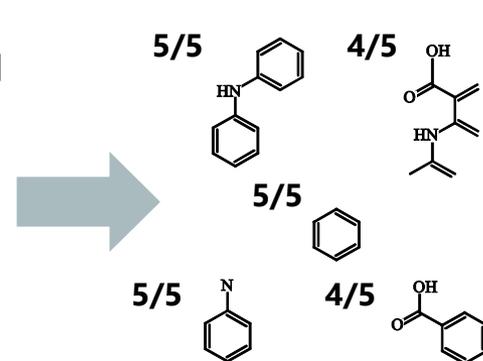
パターン要約

- ◆ Wildcardを許容した頻出飽和・極大部分グラフ

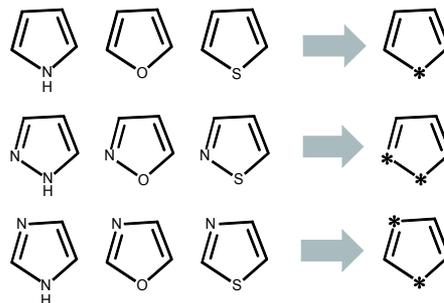
グラフデータベース



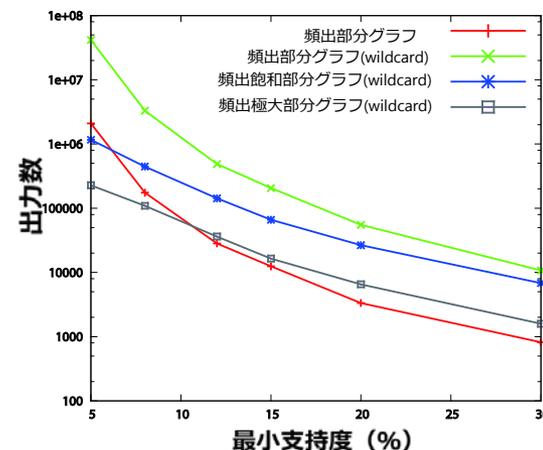
頻出部分グラフパターン



一部頂点ラベルが異なる部分構造 例) ヘテロ五員環のwildcardパターン



実験結果



- 集団としての運転行動の分析: 運転支援システムへの応用
- 交通流を表すセルオートマトンによるドライバーモデリング

セルオートマトン(CA)の統計的表現

混合二項分布



LDAによる拡張



データセット = 文書とみなす

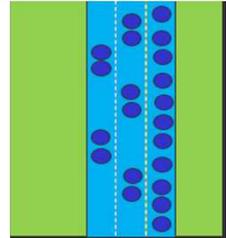
分析が容易なCA

データセット毎の特徴をとらえた
コンパクトな表現が可能

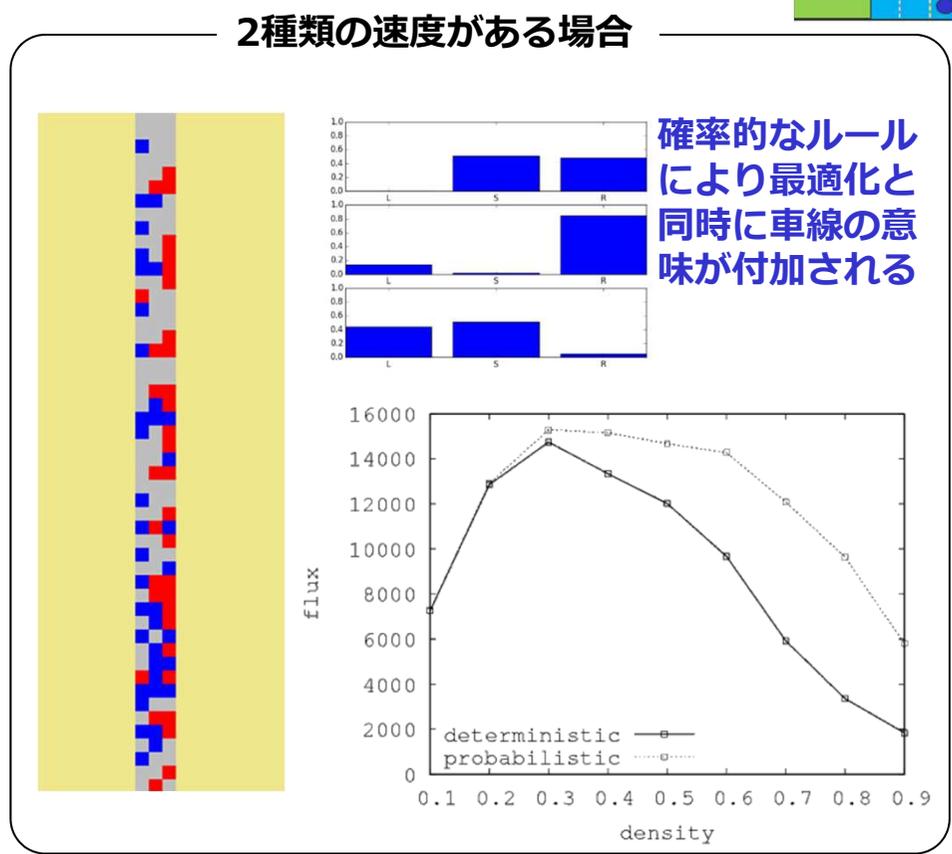
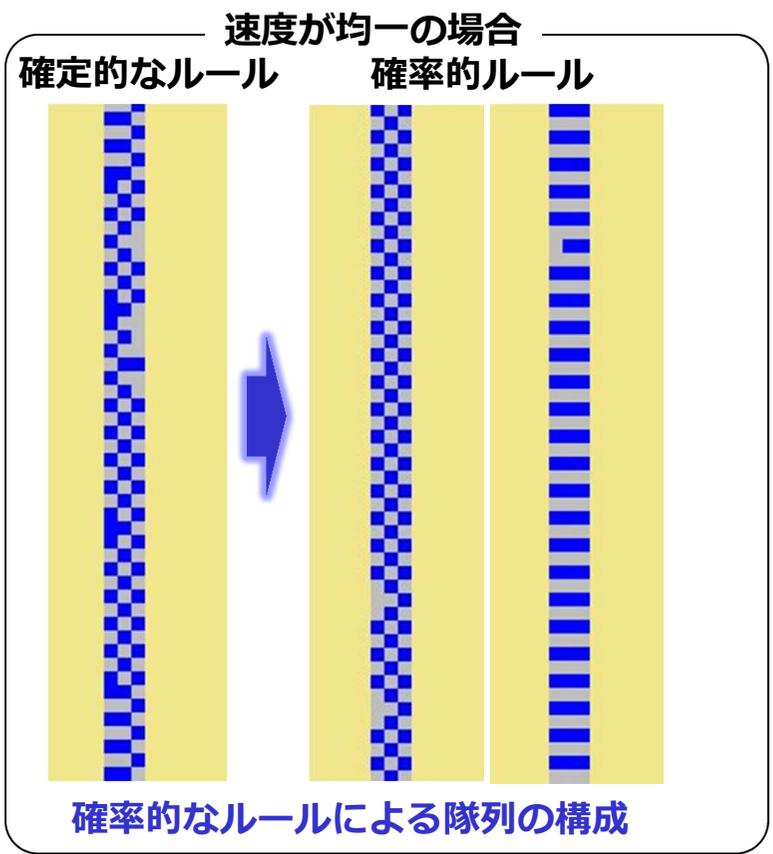
ベイズ推定によるトピックモデル

T-07 セルオートマトンを用いた自律的な車線変更モデルによる交通流最適化の研究

梶 大介 (デンソー) ・ 山崎啓介 (東工大)



自動運転車が普及した社会を想定し、3車線道路における自律的な車線変更モデルによる交通流の最適化をセルオートマトンを用いて検討

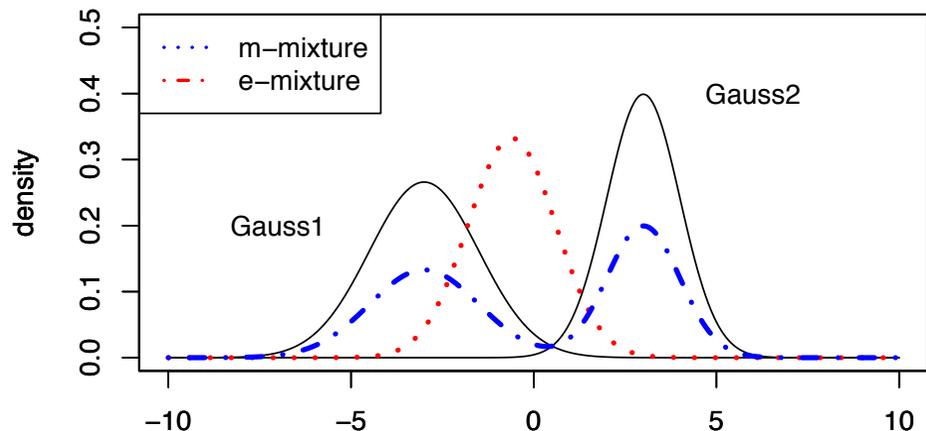


いずれの場合においても確定的なルールのみでなく、確率的なルールを与えることが車線の有効利用、交通流最適化に有効であることが示された。

T-08:ノンパラメトリックモデルのe混合推定とその応用

高野 健 (発表者・早大) ・日野英逸 (筑波大) ・赤穂昭太郎 (産総研) ・村田 昇 (早大)

情報幾何ではe混合とm混合という2つの混合モデルを考えることができる。



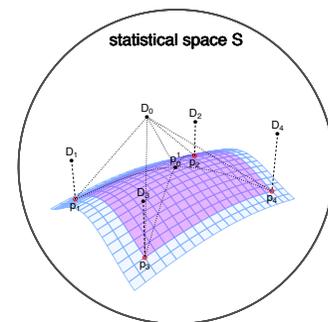
$$m\text{-mixture: } p^m = \sum_{i=1}^N \theta_i p_i$$

$$e\text{-mixture: } p^e = \exp \left(\sum_{i=1}^N \theta_i \log p_i - b(\theta) \right)$$

$$p^e(x; \theta) = \exp \left\{ \sum_{i=1}^N \theta_i \log \frac{1}{n_i} \sum_{k=1}^{n_i} \delta(x - x_k^{(i)}) - b(\theta) \right\}$$
$$\simeq \sum_{k=1}^K w_k \delta(x - y_k)$$



- 1.幾何学的な観点からアルゴリズムを構成
- 2.転移学習のようなアプローチで応用



ノンパラメトリックモデルのe混合

学生優秀プレゼンテーション賞対象

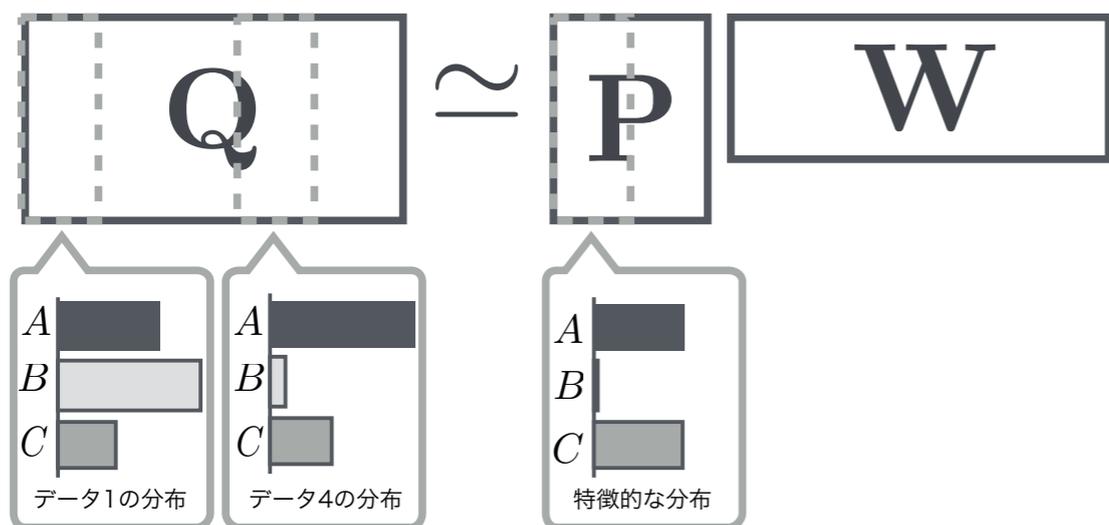
T-09: 非負値行列分解の情報幾何

奈良 寧々花¹・高野 健¹・日野 英逸²・赤穂 昭太郎³・村田 昇¹ 学生優秀プレゼンテーション賞対象

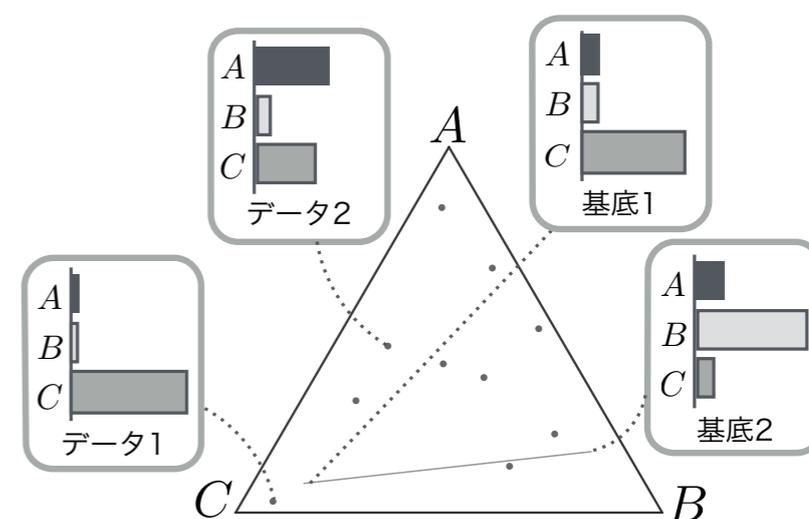
[1]早稲田大学大学院先進理工学研究科 [2]筑波大学システム情報工学研究科 [3]産業技術総合研究所

情報幾何の観点から、新しいNMFのアルゴリズムを構成する

行列分解 $X \simeq DC$ における各行列に
列和を1とする制約を入れた行列 Q , P , W を作る。



確率密度関数の空間で
データを扱うことが可能となる。

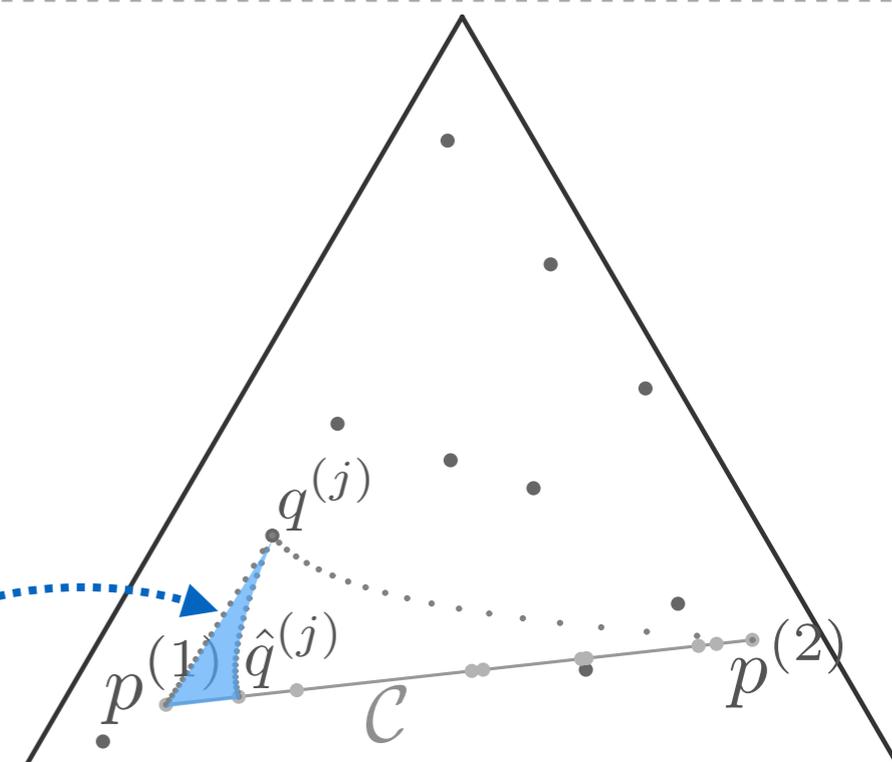


確率密度関数の空間で拡張ピタゴラスの定理を
成り立たせることで、間接的に目的関数である
近似値 $\hat{q}^{(j)}$ とデータ点 $q^{(j)}$ とのKL距離

$$D_{KL}(\hat{q}^{(j)}, q^{(j)})$$

の最小化を行うことができる。

拡張ピタゴラスの定理が
成り立つような $\hat{q}^{(j)}$ を求める。



T-11: τ -情報幾何における q -正規分布族の扱いについて



応用数学科
Applied Mathematics

福岡大学 田中 勝

- 互いに絶対連続な測度の集合上に平行移動を定義しアファイン空間にする
- τ -アファイン共役を定義する
- 縮約を定義する
- くり込みを定義する



τ -情報幾何学

非指数型分布族である q -正規分布族を用いて具体的にどうなるのかを示す

τ -情報幾何学を使う利点の1つ:

離散分布に対するエントロピーから連続分布に対するエントロピーを導出する

発散項が出てくる  くり込みにより解決される

ランダムウォークグラフカーネルの停止に関する解析 (T-12)

杉山 磨人 (大阪大学) Karsten Borgwardt (ETH Zürich)

グラフ G と G' の間の k ステップランダムウォークカーネル

$$K_{\times}^k(G, G') = \sum_{i,j=1}^{|V_{\times}|} \left[\sum_{l=0}^k \lambda_l A'_{\times}{}^l \right]_{ij} \quad (\lambda_l > 0)$$

は幾何ランダムウォークカーネル

$$K_{\text{GR}}(G, G') = \sum_{i,j=1}^{|V_{\times}|} \left[\sum_{l=0}^{\infty} \lambda^l A'_{\times}{}^l \right]_{ij} = \sum_{i,j=1}^{|V_{\times}|} \left[(\mathbf{I} - \lambda A_{\times})^{-1} \right]_{ij}$$

よりもベースラインに適している

自己情報コントロール設定における 線形分類学習の理論的脆弱性評価

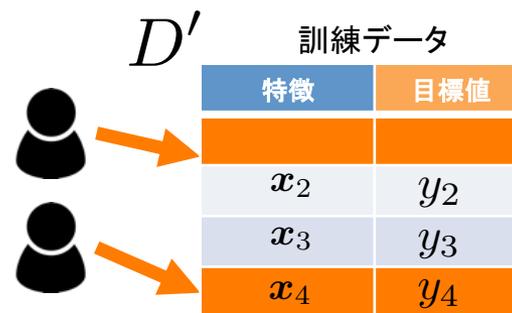
小林星平(筑波大), 奥村翔太(名工大), 竹内一郎(名工大), 佐久間淳(筑波大)

- 自己情報コントロール: 提供者自身が自分に関する情報をコントロールするという積極的・能動的なプライバシー保護

D 訓練データ

特徴	目標値
x_1	y_1
x_2	y_2
x_3	y_3

$$w = \mathcal{A}(D)$$



自己情報コントロール $w' = \mathcal{A}(D')$

提供者が自分のデータを追加・変更・削除

脆弱性: データの追加・変更・削除によって起こる汎化損失の変化

$$V_{\mathcal{A}}(w', w) := \mathbb{E}_{(x, y) \sim P} [\ell(x^{\top} w', y) - \ell(x^{\top} w, y)]$$

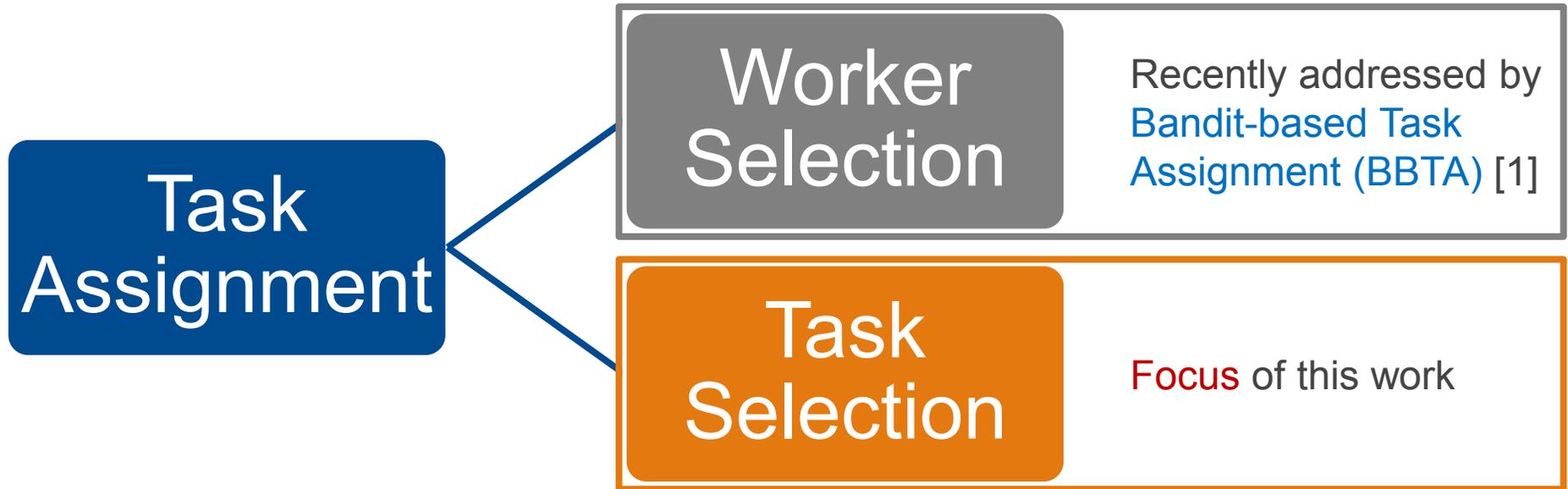
- 貢献: 既存研究ではできなかった, 訓練データの追加・変更・削除すべてに対応する理論的脆弱性評価を実現
 - 操作されるデータ数を m , 訓練データ数を n とするとデータの追加・変更・削除に対するロジスティック回帰・SVM の脆弱性は $V_{\mathcal{A}}(w', w) = O(m/n)$

T-14 Task Selection for Bandit-Based Task Assignment in Heterogeneous Crowdsourcing

Hao Zhang (Tokyo Tech) and Masashi Sugiyama (UTokyo)

Task Assignment

Online decision making on **which labeling task** is to be assigned to **which worker**



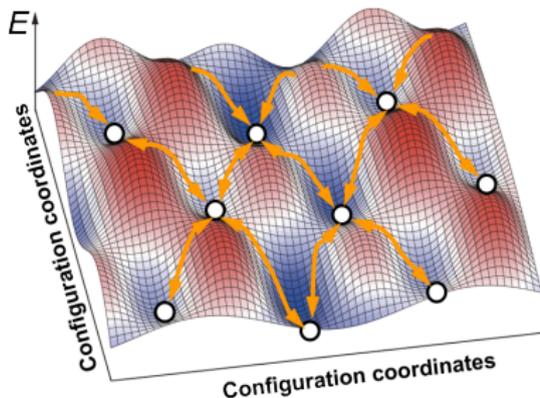
We borrowed ideas from active learning and developed **task selection** strategies for task assignment in crowdsourcing.

[1] Zhang H., Ma Y., and Sugiyama M. **Bandit-based task assignment for heterogeneous crowdsourcing**. *Neural Computation*, vol.27, no.11, pp.2447-2475, 2015.

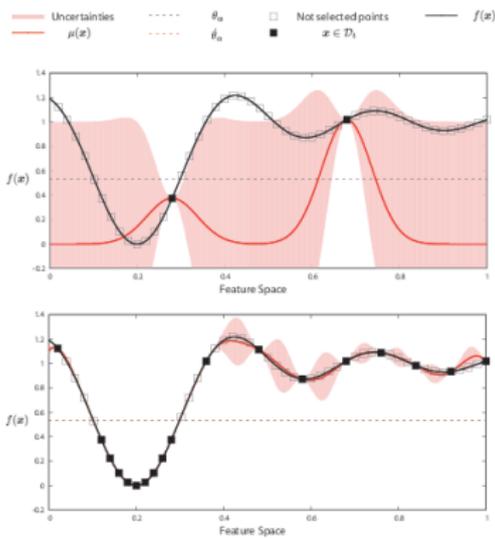
(T-15) ガウス過程を用いた選択的サンプリングとその応用

平野大輔 (名工大), 豊浦和明 (名大), 世古敦人 (京大), 志賀元紀 (岐阜大),
桑原彰秀 (JFCC), 烏山昌幸 (名工大), 設楽一希 (京大), 竹内一郎 (名工大)

- ▶ 目的: 未知な関数上の興味ある領域を効率的にサンプリング
- ▶ 方法: ベイズ最適化を拡張
 - ▶ 閾値の推定と未計算点の割合算出



電子の移動経路



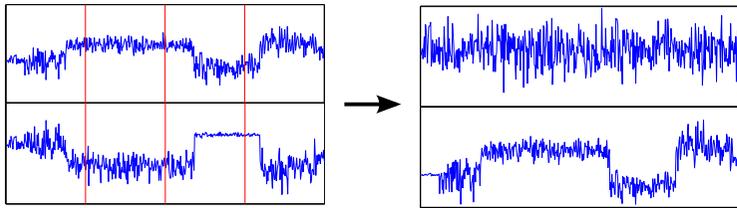
提案法

Geometry-aware Stationary Subspace Analysis for Multivariate Signals

Inbal Horev (TokyoTech), Florian Yger (Universite Paris-Dauphine) and Masashi Sugiyama (UTokyo)

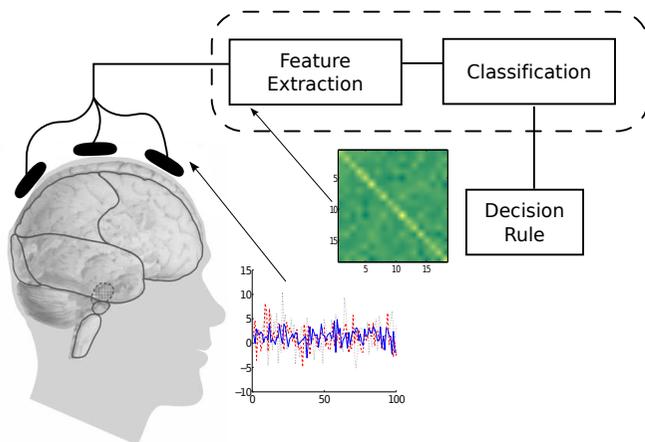
Stationary Subspace Analysis (SSA)

extracts the stationary part of a mixture of sources



for example:

Brain Computer Interface

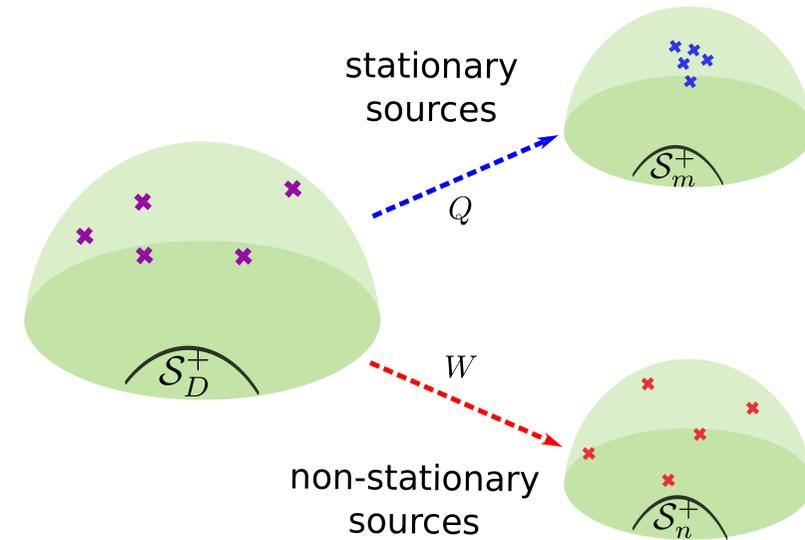
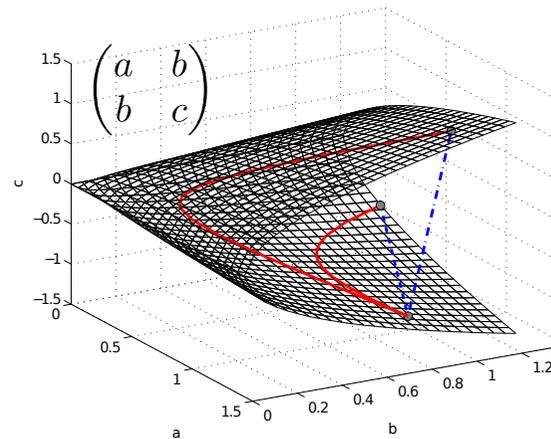


Proposal:
An SSA method based on the Riemannian geometry of SPD matrices

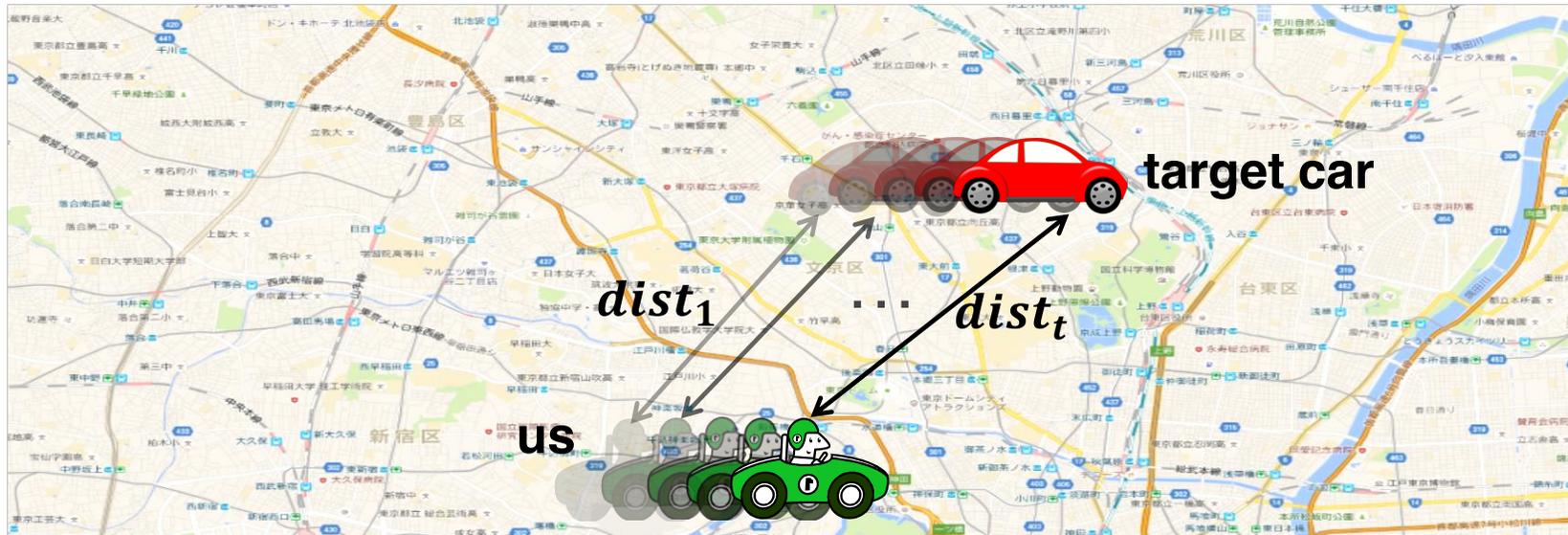
SSA: Divergence between distributions



GA-SSA: Distance between matrices



- ❑ Motivating Example: tracking a **target car** that **moves abruptly** in real time.
- ❑ Goal: Learn a time dependent driving strategy that minimizes the sum of distances between the target and us.



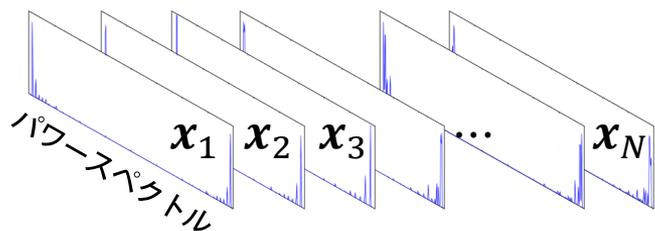
- ❑ We proposed online policy gradient algorithm for **large (continuous)** state space online MDPs with a **sublinear regret**.

音楽音響信号解析のための学生 t 分布に基づく 非負値行列分解 (NMF) と半正定値テンソル分解 (PSDTF)

吉井 和佳 (京大) 糸山 克寿 (京大) 後藤 真孝 (産総研)

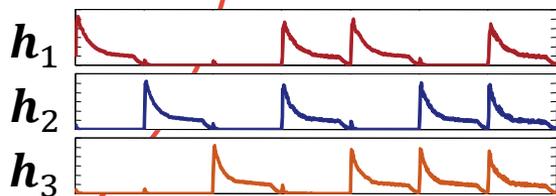
T-19

NMF: 非負値ベクトルを少数の非負値ベクトルの線形和で近似



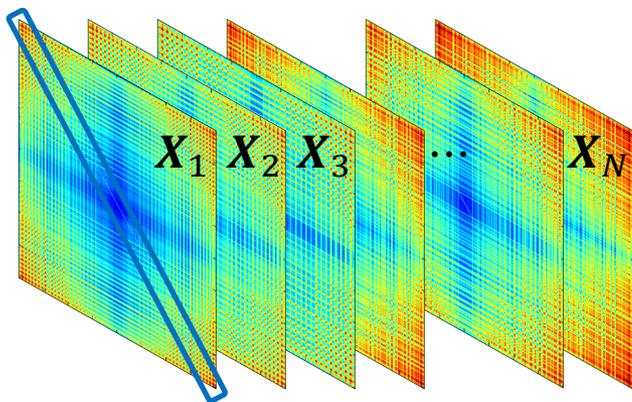
周波数ビン間の相関を考慮：位相を取り扱える！

$$\text{非負値 } x_n \approx \sum_{k=1}^K \text{非負値 } h_{kn} w_k$$

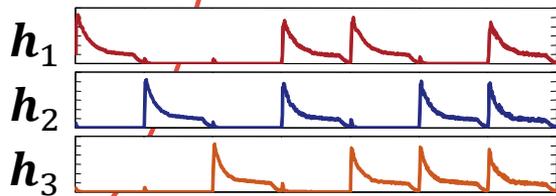


音源分離精度：SDR 18.9 [dB]

PSDTF: 半正定値行列を少数の半正定値行列の線形和で近似



$$\text{半正定値 } X_n \approx \sum_{k=1}^K \text{半正定値 } h_{kn} w_k$$



音源分離精度：SDR 26.7 [dB]

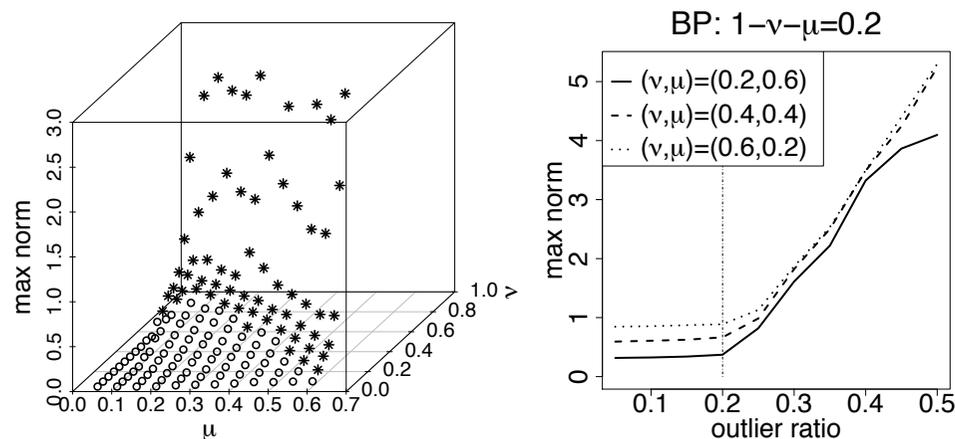
ガウス分布 \rightarrow t 分布にすると IS-NMF, Cauchy NMF, LD-PSDTF を含む 統一的な確率モデル
局所解に陥りにくい スペクトルの加法性が保証された NMF 中間的な NMF も OK

T-21 Robustification of Learning Algorithms using Hinge-loss

Takafumi Kanamori (Nagoya Univ.), Shuhei Fujiwara (Top Gate Co. Ltd.), Akiko Takeda (Univ. of Tokyo)

- **Breakdown point** : 推定法のロバスト性を測る尺度.
- ロバスト化した **SVM** の厳密な **breakdown point** を導出.
 - * **IBIS2014** : 2値判別 **SVM** の **breakdown point**
 - * **今回** :
さまざまな **SVM** の **breakdown point** (1-class, 多値, 回帰).
局所解の統計的性質を理論的に保証.

厳密な計算を可能にする解析手法 を確立



数値結果 : 理論とよく整合

L_1 正則化が不要なパラメータを削除してくれるの
ならいっその事最初にどんな分布も表せるだけのパ
ラメータを与えてみよう.

- 対象を問わずその観測データだけから自動で柔軟にモデルが作れます.
- 計算時間は2値変数20個(元モデルが 2^{20} パラメータ)のモデルで1分弱程度.
- Boltzmann machine や3体以上のポテンシャルを持つモデルも簡単に作れます.
- 隠れ変数ありでもできます.
- 実験結果はポスターで.

T-25: 予算制約つき多腕バンディット問題に対する KL-UCB の拡張戦略

渡辺 僚 中村 篤祥 工藤 峰一 (北海道大学 大学院情報科学研究科)

予算制約つき多腕バンディット問題 [Ding et al., 2013]

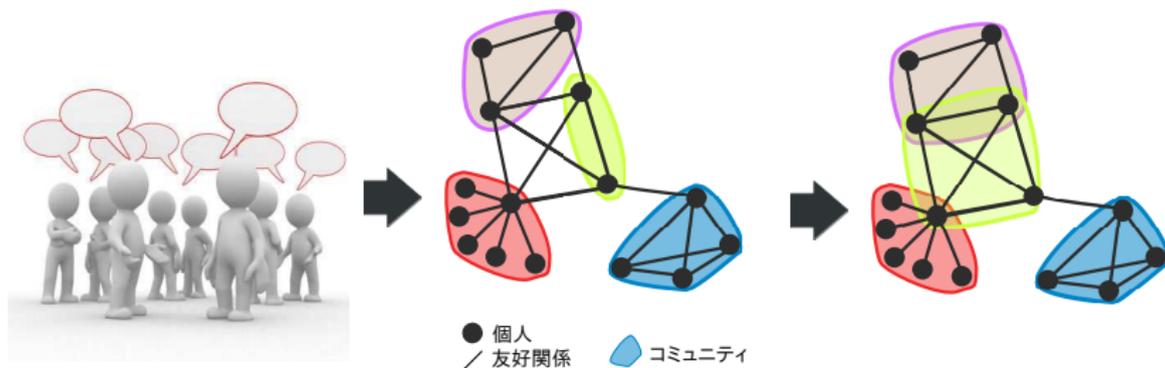
- アクション選択の度に確率的コストを負う多腕バンディット問題.
- コストは予算から支払われ, 予算が尽きるまで選択を繰り返す.
- コスト分布はアクション毎に異なる.

成果

漸近最適方策 KL-UCB [Garivier and Cappé, 2011] を予算制約つき問題に拡張した KL-UCB-IE を提案.

- 疑リグレット上界を $1/4$ 以下に改善.
- 数値シミュレーションでは TS に基づく手法 [Xia et al., 2015] と比肩するリグレット性能を達成.

T-35: 重複コミュニティ発見のための Modularity 分割の擬似クリーク拡張



- Social Network ではコミュニティの重複は自然である。
- 重複が扱えるようグラフ分割法を拡張した方法が提案されたが、一般的には構造に依存する 重複度や個数を決める必要がある。



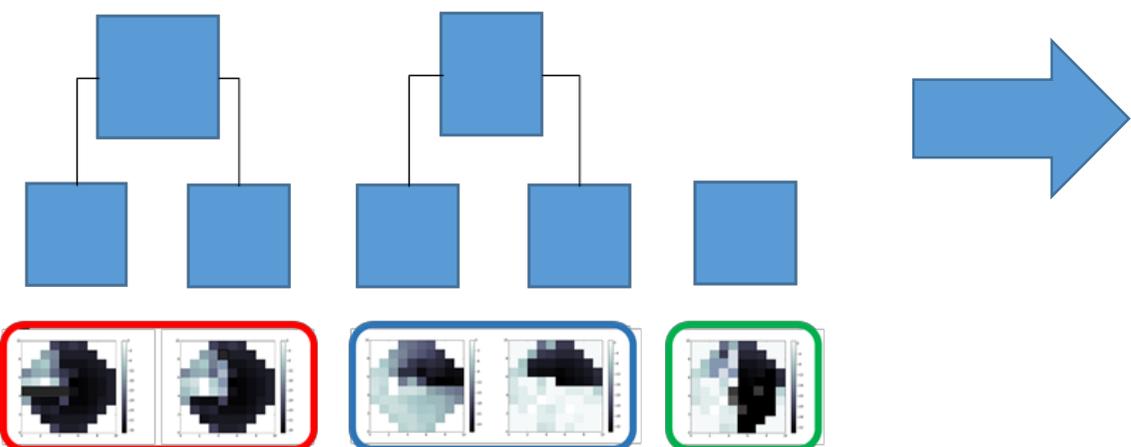
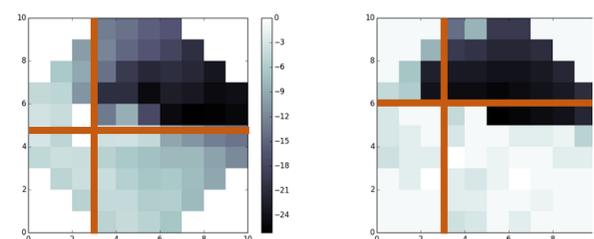
- 構造を直接決めるより本質的な拡張法を提案。
- Modularity 分割基準を拡張。他の基準の拡張も可能。

D-01 MDL規準に基づく階層的クラスタリングを用いた緑内障進行パターンに関する知識発見

○真矢 滋(東京大学*), 森野 佳生(東京大学), 村田 博史(東京大学)
朝岡 亮(東京大学), 山西 健司(東京大学, CREST)
(*現在 (株)東芝 研究開発センターに所属)

- 緑内障とは
 - 症状: 視野欠損
 - 多様な進行パターンが存在
- 提案手法 [1]
 - 直積構造を組み込みクラスタ数自動決定が可能なMDL規準による階層的クラスタリング
 - 動的計画法を用いた効率的な符号長算出

緑内障の特徴: 直積構造



医学的な知識の獲得

1. 緑内障進行の時間発展を検出
2. 緑内障進行予測の精度改善に貢献

[1] S. Maya, K. Morino, H. Murata, R. Asaoka, and K. Yamanishi. "Discovery of glaucoma progressive patterns using hierarchical MDL-based clustering," In *Proceedings of the 21th ACM SIGKDD*, pp. 1979-1988, 2015.

D-03 スパース正則化に対するDCアルゴリズム

後藤 順哉 (中央大学) ・ 武田 朗子 (東京大学) ・ 東野 克哉 (東京大学)

L0制約と等価な正則化法の提案

L0制約付き問題

$$\min. f(x)$$

$$\text{s.t. } \|x\|_0 \leq k$$

非ゼロ要素数

凸緩和

L1正則化

$$f(x) + \lambda \|x\|_1$$

上位k成分の和

$$\lambda (\|x\|_1 - \|x\|_k)$$

等価

DC正則化

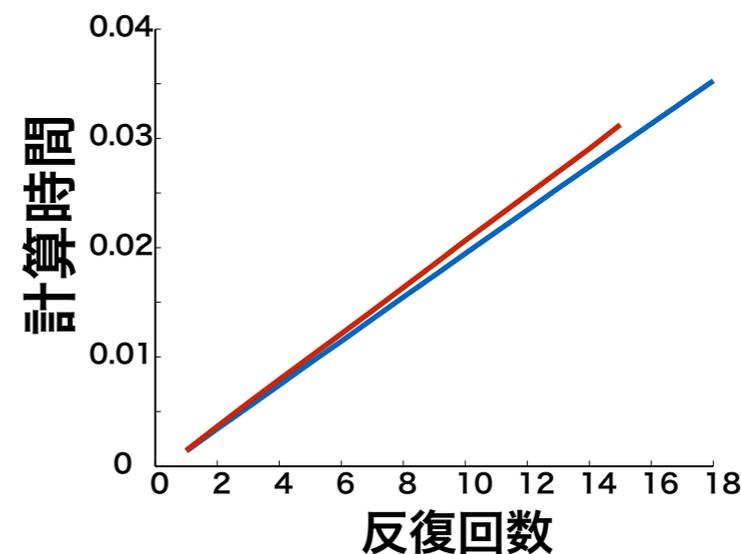
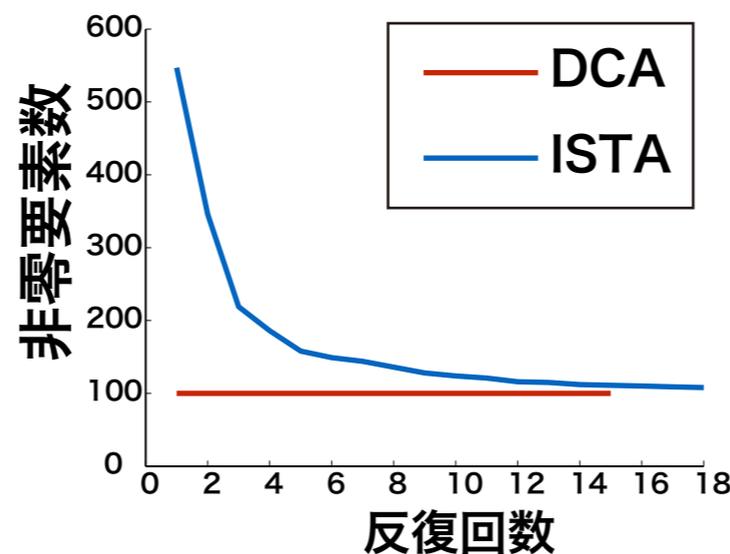
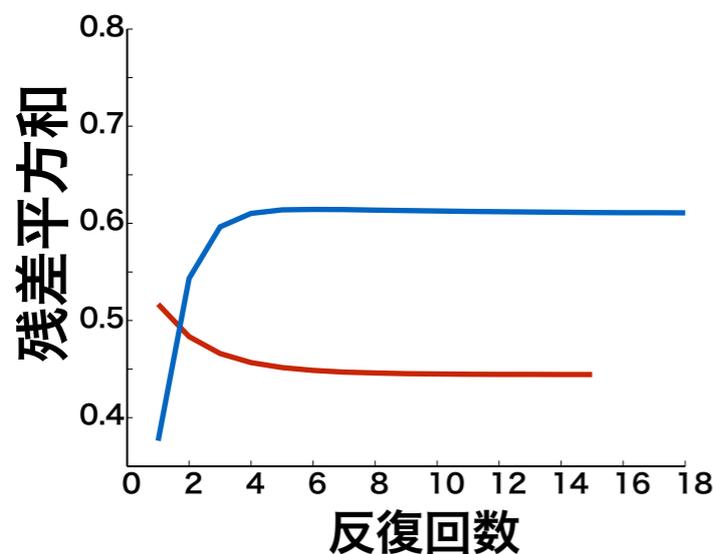
ISTA

$$x^{(t+1)} = \text{prox}_{\lambda \|\cdot\|_1} \left(x^{(t)} - \frac{1}{L} \nabla f(x^{(t)}) \right)$$

劣勾配

$$x^{(t)} - \frac{1}{L} \nabla f(x^{(t)}) + \frac{\lambda}{L} s(x^{(t)})$$

DCアルゴリズム



D-05: 高階結合正則化を用いた教師あり学習

竹内 孝 (NTT), 河原 吉伸 (阪大), 岩田 具治 (NTT)

- ✓ 特徴量のグループを用いてパラメータを結合する正則化を提案し、教師あり学習に適用
- ✓ 正則化項の最適解を $O(n \log n)$ で求める近接法 (最大流アルゴリズム) を提案
- ✓ レビューの評価点をテキストから予測するタスクにおいて予測誤差の改善を確認

高階結合正則化
(凸非平滑関数)

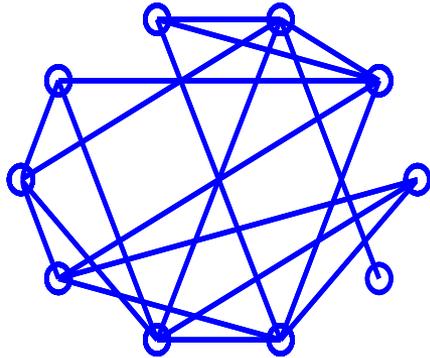
ロバスト P_n Pottsモデル
(劣モジュラ関数)

ロバッシュ拡張

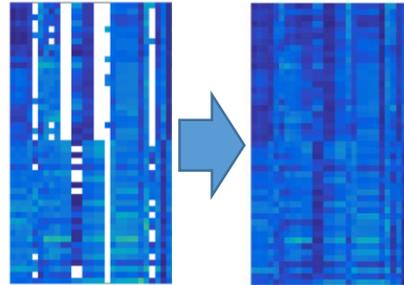
D-06 欠損ありオミクスデータのための 疎偏相関行列推定法

宇田新介, 九州大学

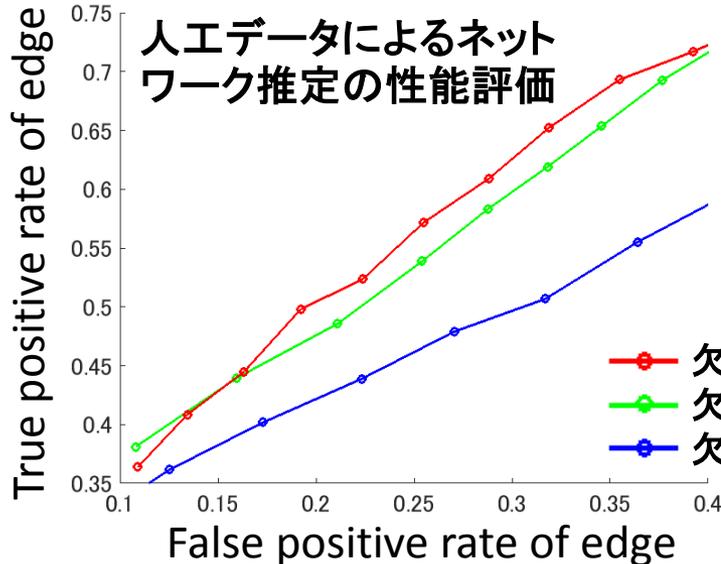
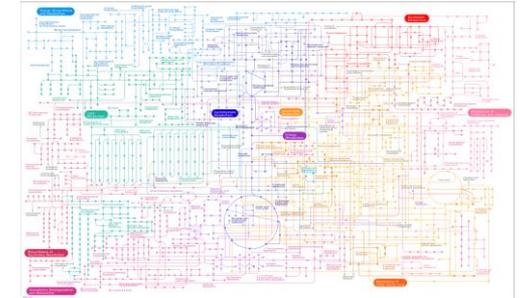
疎偏相関行列による
ネットワーク構造の予測



データ行列復元



データベース(DB)



○定式化

$$\min_{\mathbf{p}, \mathbf{q}, \rho, \sigma} \left\{ \sum_{i,j \in \mathbf{D}} \frac{1}{2} \left(y_{ij} - \sum_h p_{ih} q_{jh} \right)^2 + \frac{\beta}{2} \sum_{i,j} w_j \left(\sum_h p_{ih} q_{jh} - \sum_{k \neq i} \rho^{ik} \sqrt{\frac{\sigma^{kk}}{\sigma^{ii}}} \sum_h p_{ih} q_{jh} \right)^2 + \sum_{i>j} \gamma_{ij} |\rho^{ij}| \right\}$$

L1正則化 + 行列積分解低ランク近似

- 欠損, 偏相関同時, DBあり (提案手法) ※ 疎偏相関行列とデータ行列の復元を同時に行う
- 欠損, 偏相関個別, DBあり
- 欠損, 偏相関個別, DBなし

ロバストなスパース回帰 バッチおよびオンライン学習

川島 孝行(総研大) 藤澤 洋徳(統数研)

Lasso

$$\begin{aligned}\hat{\beta} &= \arg \min_{\beta} \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_{i1} \cdots - \beta_p x_{ip})^2 + \lambda \sum_{j=1}^p |\beta_j| \\ &= \arg \min_{\beta} D_{KL}(\bar{g}(x, y), f(x, y; \beta)) + \lambda \sum_{j=1}^p |\beta_j|\end{aligned}$$

標準化 + Lasso

$$\begin{aligned}\hat{\beta} &= \arg \min_{\beta} \sum_{i=1}^n (y_i - \beta_0 - \beta_1 \frac{x_{i1} - \bar{x}_1}{S_1} \cdots - \beta_p \frac{x_{ip} - \bar{x}_p}{S_p})^2 + \lambda \sum_{j=1}^p |\beta_j| \\ &= \arg \min_{\beta} D_{KL}(\bar{g}(x, y), f(y|x; \beta) \prod_{j=1}^p f(x_j; \mu_j, \sigma_k^2)) + \lambda \sum_{j=1}^p |\beta_j|\end{aligned}$$

標準化のダイバージェンスを用いた表現.

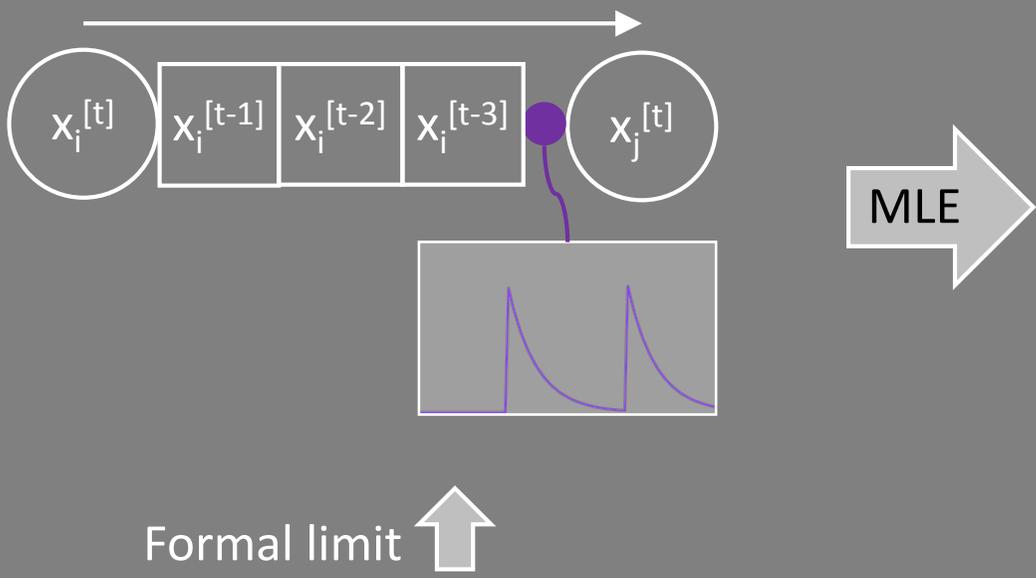
提案手法

γ ダイバージェンスによるロバスト化.

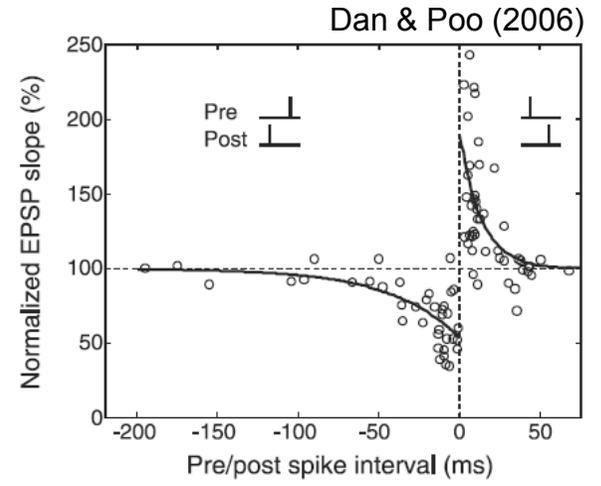
- ロバストな推定アルゴリズム
 - ➡ MMアルゴリズムを用いた, 単調な推定アルゴリズム.
- 新規データに対するスケール不変性
 - ➡ オンライン最適化による, 逐次推定アルゴリズム.
- 標準化のロバスト化
 - ➡ 説明変数に対するロバスト性.

Osogami & Otsuka (2015) www.nature.com/articles/srep14149

Dynamic Boltzmann machine

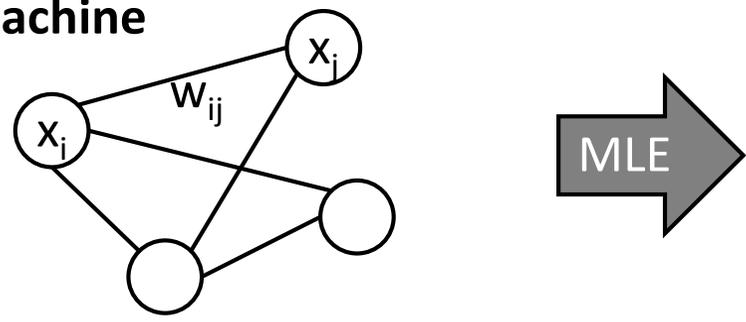


Spike-timing dependent plasticity



↑ Refine

Boltzmann machine



Hebb's rule

Cells that fire together, wire together

$$w_{ij} \leftarrow w_{ij} + \eta(x_i x_j - \langle x_i x_j \rangle_w)$$

D-10 相対評価に基づく協調ランキング問題

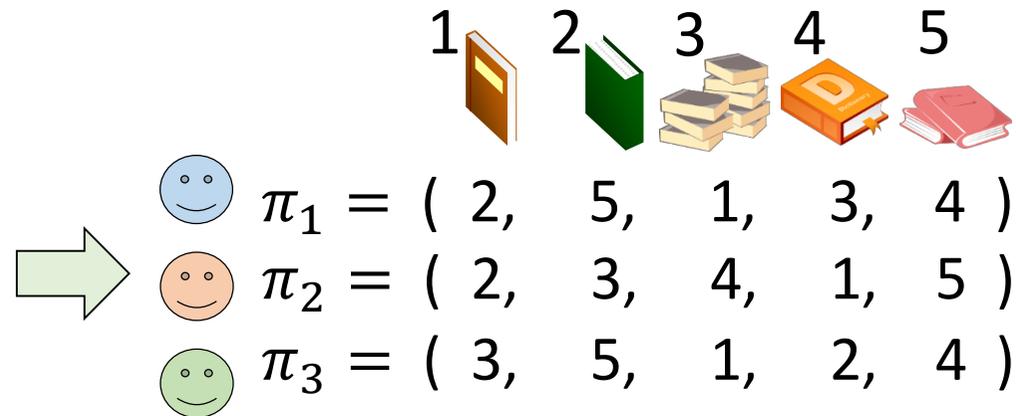
森富賢一郎・畑埜 晃平・瀧本英二(九州大学)

協調ランキング問題

アイテムペアの比較評価情報

	1 2	1 3	...	3 4	3 5	4 5
Y	 	 	...	 	 	 
	?	1	...	0	?	0
	0	?	...	1	?	1
	?	1	...	0	1	?

アイテム上のランキング



提案手法

一部のアイテムペアの比較評価情報

絶対評価 \hat{X}

緩和した相対評価 \hat{Y}

緩和した
相対評価 \hat{Y}

対応するランキングにラウンディング

相対評価情報を連続値に緩和

- 潜在的な絶対評価を考えるモデル
- 相対評価の穴埋めをするモデル

主結果

ランキングの汎化誤差の上界の導出

学生優秀プレゼンテーション賞対象

D-11

低ランク分解を用いた ノンパラメトリックテンソル回帰

今泉允聡 (東京大学) 林浩平 (情報学研究所)

テンソル入力のノンパラ回帰

- ・ 入力テンソルの低ランク分解
- ・ ノンパラ回帰関数の分解



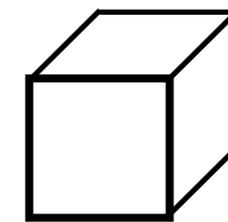
推定誤差の改善

Nonparametric regression

$$Y = f(X) + \epsilon$$

Tensor data

$$X \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_K}$$



3-mode
tensor

具体的な操作

$$X \mapsto \{x_r^{(k)} \in \mathbb{R}^{I_k}\}_{k,r} \mapsto \sum_m \sum_r \prod_k f_m^{(k)}(x_r^{(k)})$$

Tensor input

CP-decomposition

Regression model

D-13

Local Kernel Dimensional Reduction in Approximate Bayesian Computation

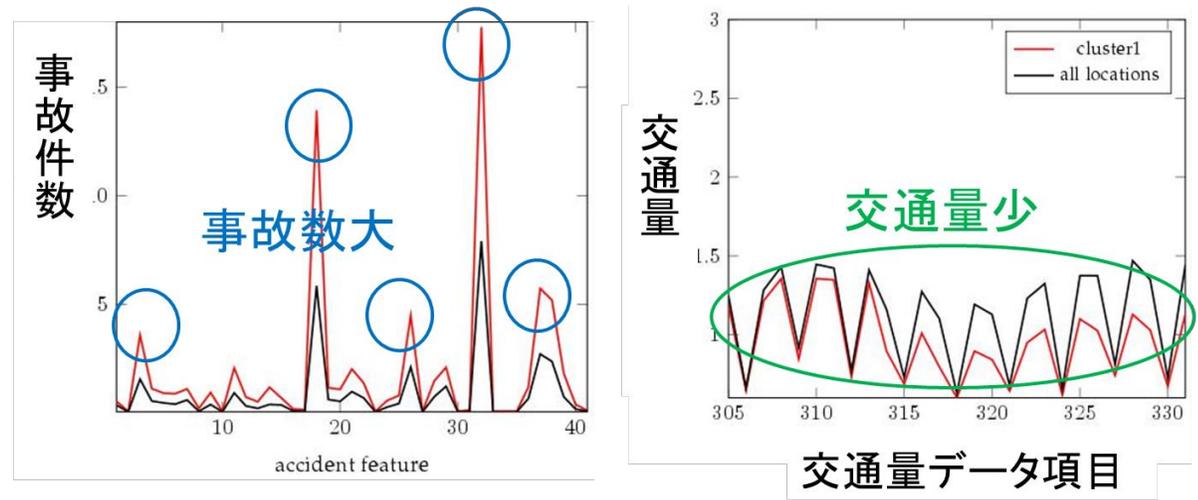
周晋(SOKENDAI), 福水健次(ISM)

目的	Automatically construct sufficient summary statistics in Approximate Bayesian Computation like algorithms.
方法	<ul style="list-style-type: none">• Using Gradient based Kernel Dimensional Reduction to reduce the dimensionality of the summary statistics.• Propose a Local weighting algorithm to concentrate on the observation.• Different summary statistics for different parameter.
結果	<ul style="list-style-type: none">• Low dimensionality achieved.• Accurate estimation of the posterior mean measured in mean squared error.• Works well in highly nonlinear problems.

D-15: 特徴ベース非負値行列分解に基づく交通リスクマイニング

発表者: 守屋 航一・松島慎・山西健司 (東京大学大学院情報理工学系研究科)

- ヘテロな交通データから道路地点の危険度を予測する問題
- 「**特徴ベース非負値行列分解**」という新しい手法を提案
- これを用いた危険箇所のクラスタリング・事故件数の予測を行うフレームワークを提案
- 提案手法は既存手法より高速かつ高精度であることを実験で確認
- 実際の交通データに対して用いた結果、**危険箇所の特定に成功**



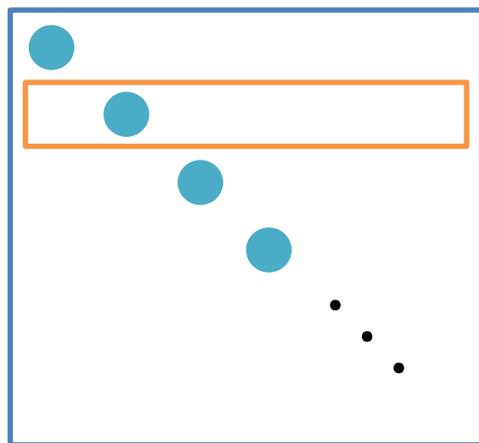
赤: クラスタ1の平均ベクトル
黒: 全体の平均ベクトル

IEEE DSAA' 2015 (Paris)で発表
本研究はJST-CRESTの一部として
行われた

類似度に基づくクラスタリング

竹岡 邦紘 岡留 剛（関西学院大）

クラス数を未知とした カーネル行列に対するクラスタリング手法の提案



カーネル行列

ポイント

- 確率的生成モデル
行列の各要素が潜在特徴ベクトル \mathbf{z}_i の内積で生成
- 行列の各要素が非独立のため、最適化が困難
→ 行列を行ごとに分解し、行ごとの尤度を求める
 - ・ 対角要素が与えられた下で、他の要素が独立
 - ・ 尤度は \mathbf{z}_i の次元数によらない
- 「事後確率最大」で、分類しクラス数も推定

活性値情報のグループ化と ランク学習による活性化化合物予測

D-19

学生優秀プレゼンテーション賞対象

鈴木 翔吾, 大上 雅史, 秋山 泰 (東京工業大学 大学院情報理工学研究科 計算工学専攻)

目的

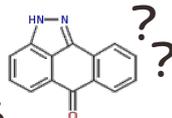
活性が未知の化合物のランキングを予測する

➡ 既知の活性値情報を用いたランク学習

問題

化合物の活性値情報には**実験誤差**が潜在する

- 実験系による誤差
- 実験者・実験手法によっても誤差が生じる
- オーダーレベルの誤差が生じることもある



化合物	A	B	C	D
真の活性値	10 nM	300 nM	500 nM	1,000 nM
観測値	15 nM	600 nM	450 nM	10,000 nM

実験誤差

観測値のランクは容易に入れ替わる

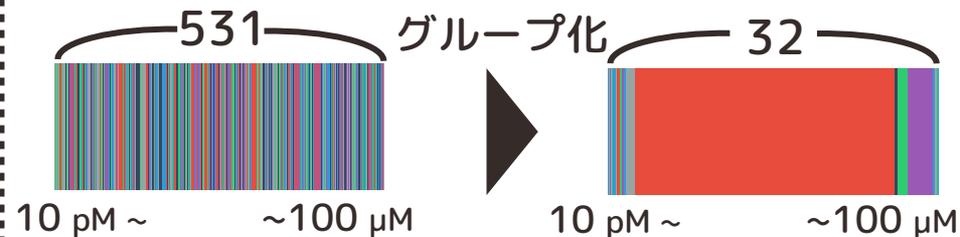
➡ 細かいランクに意味は無い
(特に活性の有無が曖昧な領域)

提案手法

観測値によって化合物をグループ化する

➡ グループランクを用いたランク学習

化合物	A	B	C	D
観測値	15 nM	600 nM	450 nM	10,000 nM
グループ ランク	1	2	2	3

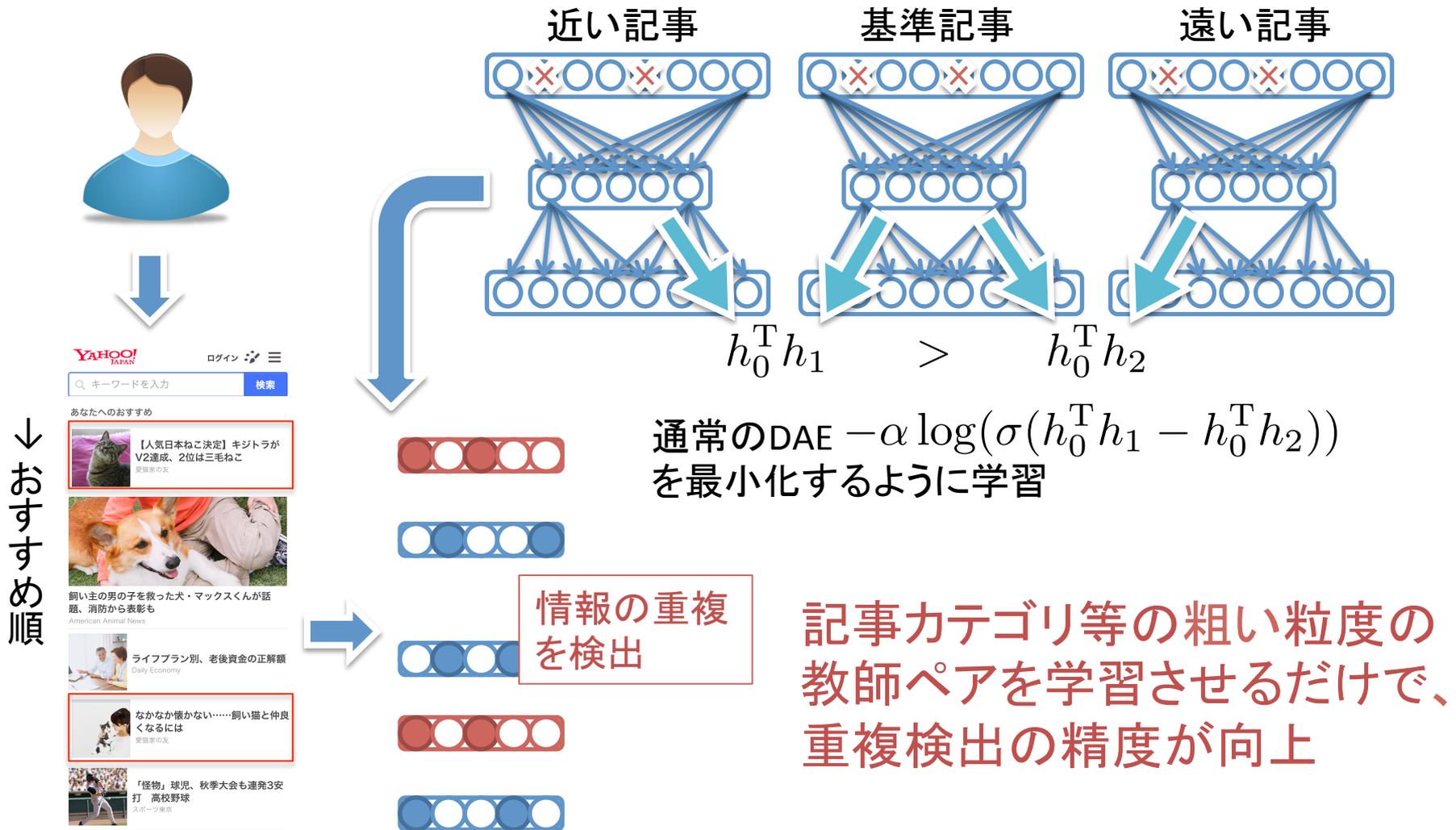


	従来手法	提案手法
ROC-AUC	0.846	0.862
NDCG@10	0.289	0.341

※データセット: BindingDB CTSK

ニュースレコメンドにおける 分散表現を用いた重複記事排除

大倉 俊平, 田頭 幸浩, 田島 玲(ヤフー株式会社)



ネットワーク中心性を用いた テンポラルネットワークの異常検知

○要名本義太郎(東大), 森野佳生(東大), 山西健司(東大, CREST)

研究背景

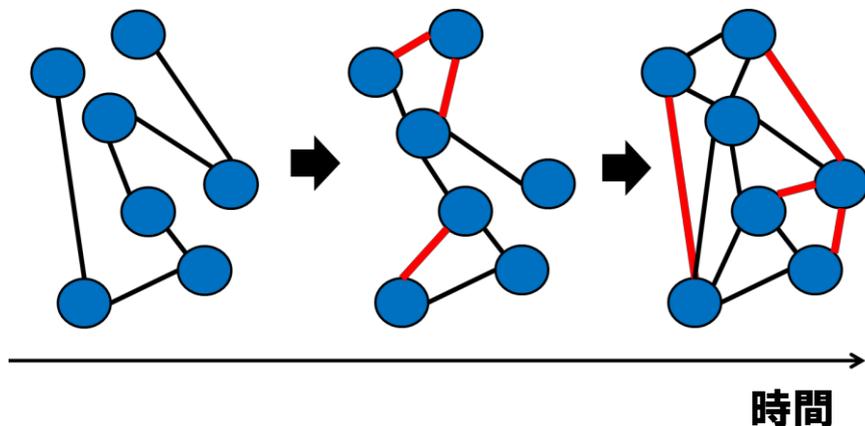
- ネットワークの重要な指標として、**ネットワーク中心性**がある
- 近年**テンポラルネットワーク**の異常検知が盛んだが、ネットワーク中心性の利用については十分に解析されていない

研究目標

- ネットワーク異常検知におけるネットワーク中心性という指標の効果・役割について明らかにする
- 複数の**ネットワーク中心性を組み合わせることでネットワークの異常検知精度を向上させる

提案手法

- 複数のネットワーク中心性を組み込んだ、テンポラルネットワークの異常検知手法
- 既存手法 (Panagiotis+, 2010) より**異常検知精度が向上**



D-27: fMRIデータに対するシンプルで強い仮定を必要

としない脳活動領域の特定法

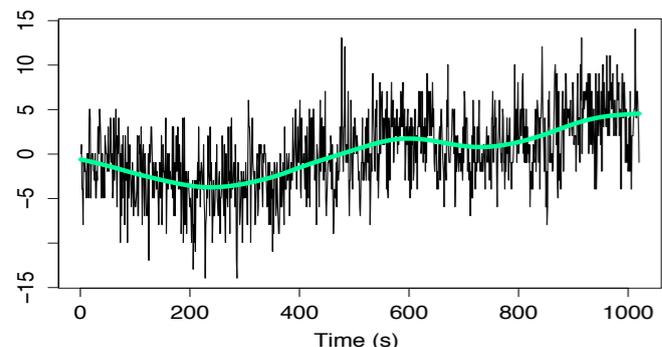
寺田 吉壱 /



📌 代表的な解析ソフト(SPM等) ⇒ ドリフトの完全除去や正規性を仮定

★ 弱い仮定の下, simpleなHRF推定量 (DBE) の一緻性と漸近正規性を証明

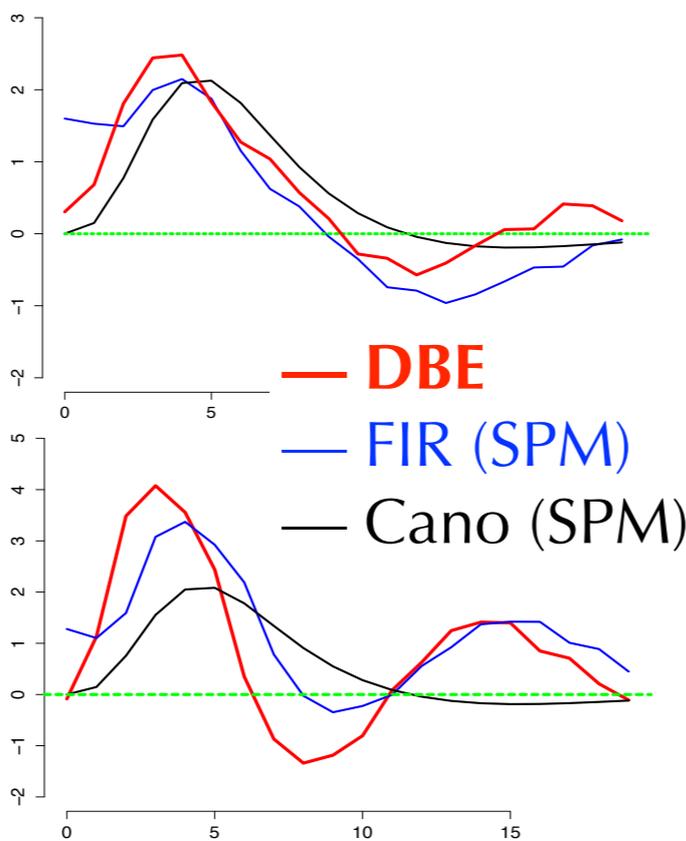
★ この結果に基づき活動領域特定のための新しい検定統計量を提案!



← あるvoxelの実際のfMRI signal
(装置等の影響による信号のドリフト)

Remark :
脳活動 \Rightarrow HRF $\neq 0$

血流動態反応関数 (HRF) の推定結果の比較

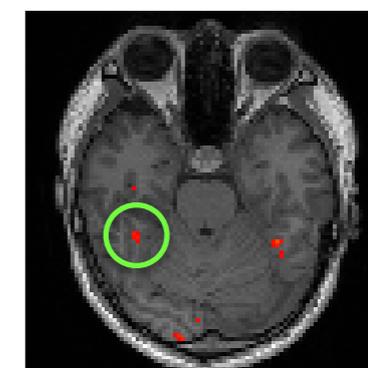


Canonical (SPM)



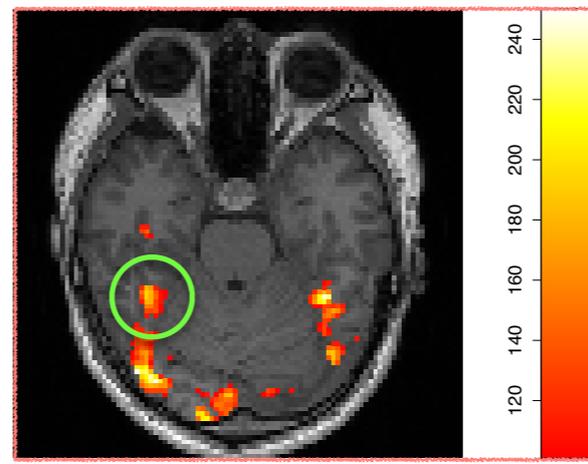
提案手法

FIR (SPM)

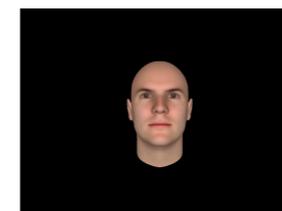


← 従来法 (SPM)

FFA (Fusiform face area) の活動がより明確に!



← 刺激

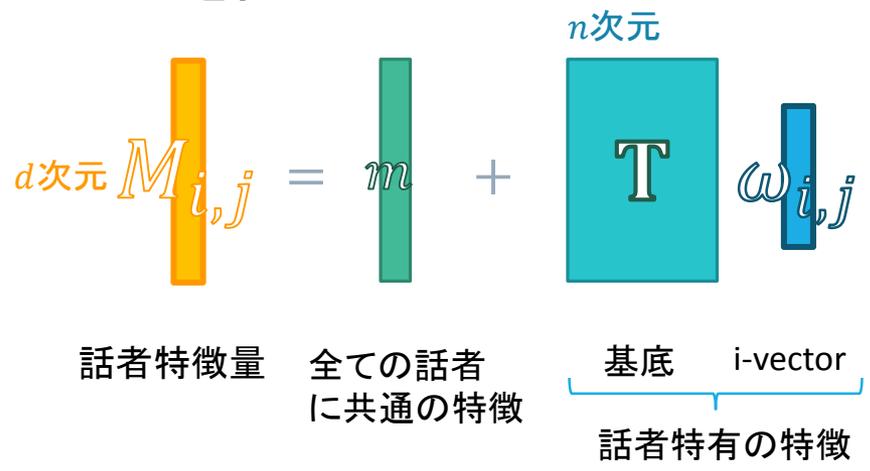


自分の脳

被験者1の脳

定式化

発話から得られた話者の特徴量から i-vectorを抽出



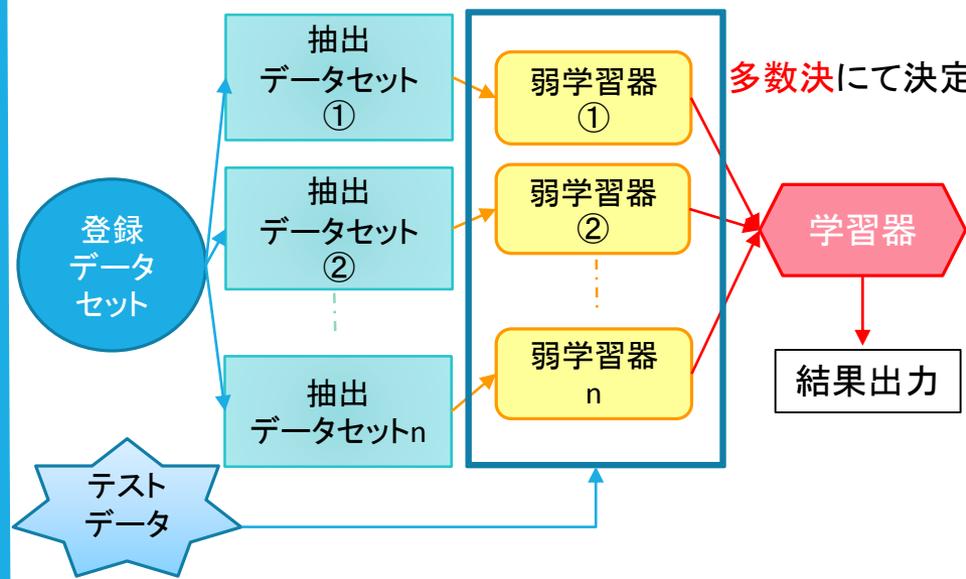
※登録データから因子分析により基底を学習

i-vectorは超球上に分布するため、cos類似度を用いて話者の推定が可能

$$\text{推定話者} : \hat{i} = \underset{i}{\operatorname{argmax}} \max_j \cos(\omega_{test}, \omega_{i,j})$$

提案手法

更なる識別精度向上のためBaggingを導入



結果

既存手法に対し、改善が見られた

D-32 オンライン広告における Extreme Multi-label分類手法の応用

田頭 幸浩（ヤフー株式会社）

- FastXMLをオンライン広告の実データに適用した
 - ラベルの種類数が極端に多い場合における
マルチラベル分類手法（Prabhu&Varma, 2014）

ユーザー ~ 100万



さまざまな広告 ~ 20万



マルチラベル分類

問題設定

連続空間上のDueling Bandit

Cost-based approach -> コスト関数に凸性を仮定

提案手法

Stochastic Mirror Descent-based Algorithm

結果

Regret の解析

- 既存のアルゴリズム $Reg_T^{DB} = O(T^{3/4})$ (凸性)
- 提案アルゴリズム $Reg_T^{DB} = \tilde{O}(T^{1/2})$ (強凸性)

凸最適化との関係

- Regret間の関係 $C_1 Reg_T^{DB} \leq Reg_t^{FO} \leq C_2 Reg_T^{DB}$
- 提案アルゴリズムは関数最適化の意味でも有効

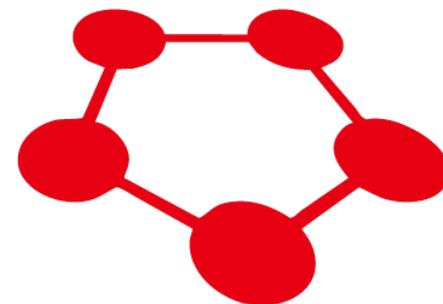
Chainer: 深層学習のための次世代オープンソースフレームワーク

D-36

得居誠也・大野健太 (Preferred Networks)
比戸将平・Justin Clayton (Preferred Networks America)

- ニューラルネットのフレームワークは多数提案されている: Caffe, Theano, Torch7, MXNet, TensorFlow, ……
- Chainer はこれらに代わるフレームワークで、異なるパラダイムにもとづく
 - 動的な計算グラフ構築
 - 分岐、ループの記述が直感的
 - デバッグしやすい
- 使い方の相談、他のフレームワークとの比較に関する議論などお待ちしております

```
# RNN example
loss = 0
for x, t in zip(xs, ts):
    h = F.tanh(
        model.x_to_h(x) +
        model.h_to_h(h))
    y = model.h_to_y(h)
    loss += loss_fun(y, t)
loss.backward()
optimizer.update()
```

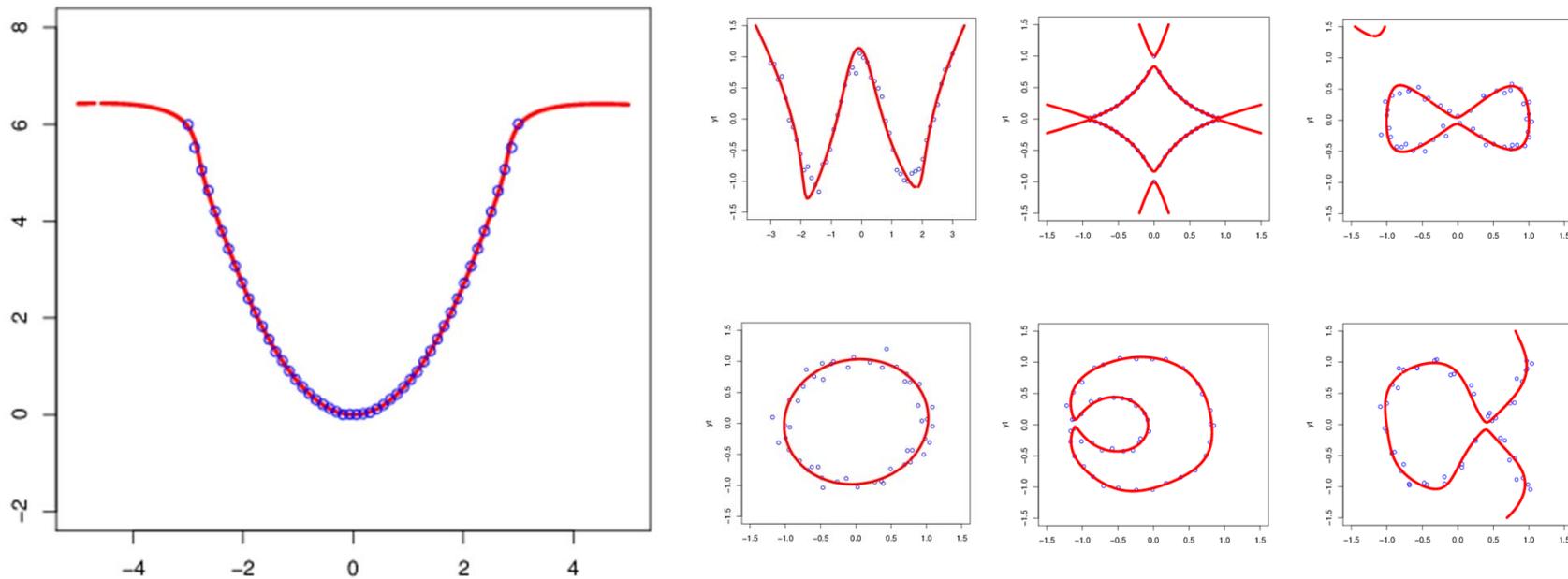




[D-38] ガウス核を用いた柔軟な超曲面あてはめに向けて

藤木 淳(福岡大学), 赤穂 昭太郎(産総研)

KMCA (核劣成分分析) により柔軟な超曲面あてはめを実現する



ガウス核では零固有値の固有空間の次元は無限大
この固有空間の自由度(自由度無限大)が柔軟なあてはめを実現する
しかし無限次元から適切なベクトルを選ぶのが難しい

特徴と標本の同時セーフスクリーニング

柴垣 篤志, 小川 晃平 (名工大), 畑埜 晃平 (九大), 竹内 一郎 (名工大)

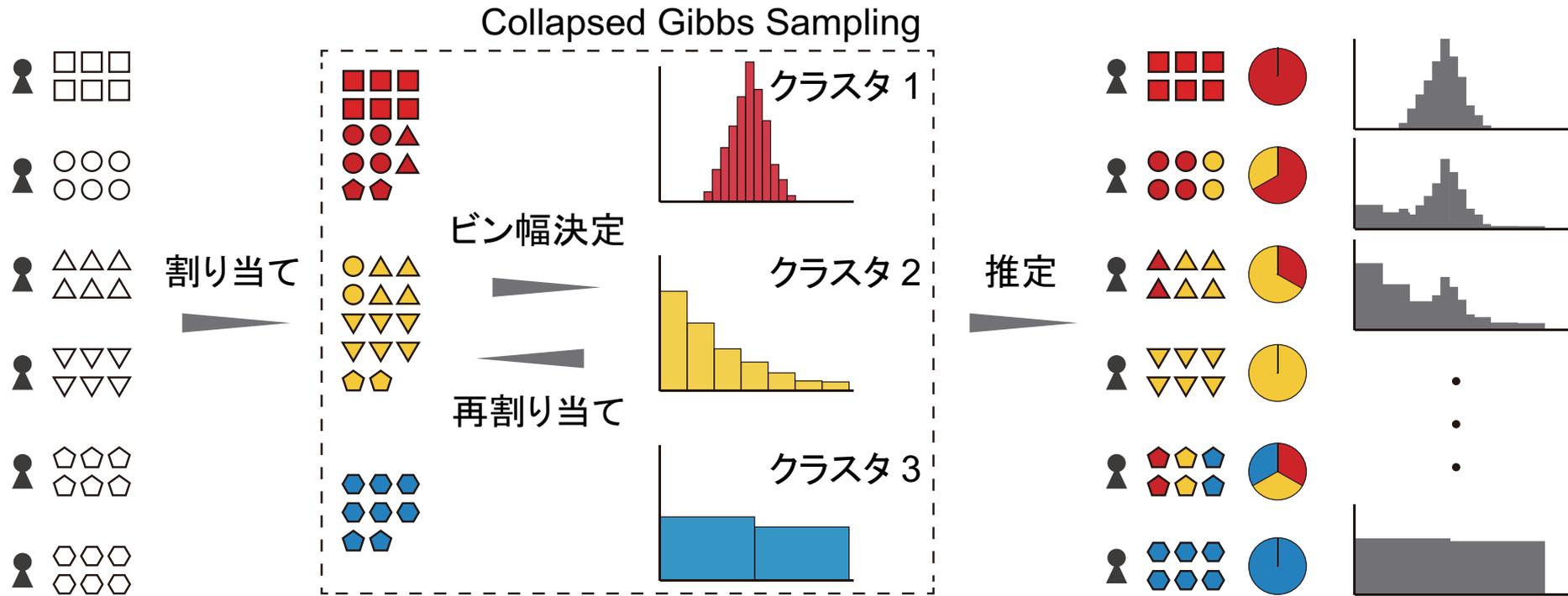
- ▶ 特徴セーフスクリーニング：学習に必要な特徴を削除
 - ▶ L1 正則化経験損失最小化学習において
- ▶ 標本セーフスクリーニング：学習に必要な標本を削除
 - ▶ L2 正則化ヒンジ損失 (SVM) において

提案：特徴と標本の同時セーフスクリーニング

- ▶ L1 正則化ヒンジ損失
- ▶ 特徴と標本を同時に考えることによる相互作用でより強いスクリーニングルールの構築

ヒストグラム密度推定法を組み込んだ 混合メンバシップモデルの提案

金 秀明, 澤田 宏 NTTサービスエボリューション研究所



- ① ノンパラな確率密度関数でクラスタを表現可能.
 - ② 簡単な Collapsed Gibbs Sampling を実装可能.
- そんな混合メンバシップモデルの提案

D-44

オンライン二分探索木問題に対する 更新コスト付きリグレット解析について

学生優秀
プレゼンテーション賞
対象

○松川 理拓, 山内 由紀子, 来嶋 秀治, 山下 雅史(九州大学)

- ◆概要：Hannanのアルゴリズム(1957)とKalai-Vempalaの遅延更新(2005)を基に改良したアルゴリズムを提案し，リグレットの上界を解析する。
- ◆貢献：入力列の長さ T を未知としても，探索コストのリグレット・探索木の更新コスト共に $o(\sqrt{T})$ を達成。

		入力列の長さ T ... 既知	入力列の長さ T ... 未知
遅延更新 なし	アルゴリズム	FPL (Kalai-Vempala)	Hannan's
	探索コストの リグレット	$o(\sqrt{T})$	$o(\sqrt{T})$
	更新コスト	$\Omega(T)$	$\Omega(T)$
遅延更新 あり	アルゴリズム	FLL (Kalai-Vempala)	本研究
	探索コストの リグレット	$o(\sqrt{T})$	$o(\sqrt{T})$
	更新コスト	$o(\sqrt{T})$	$o(\sqrt{T})$

時系列データからの 畳み込み非負値行列因子分解の構造推定

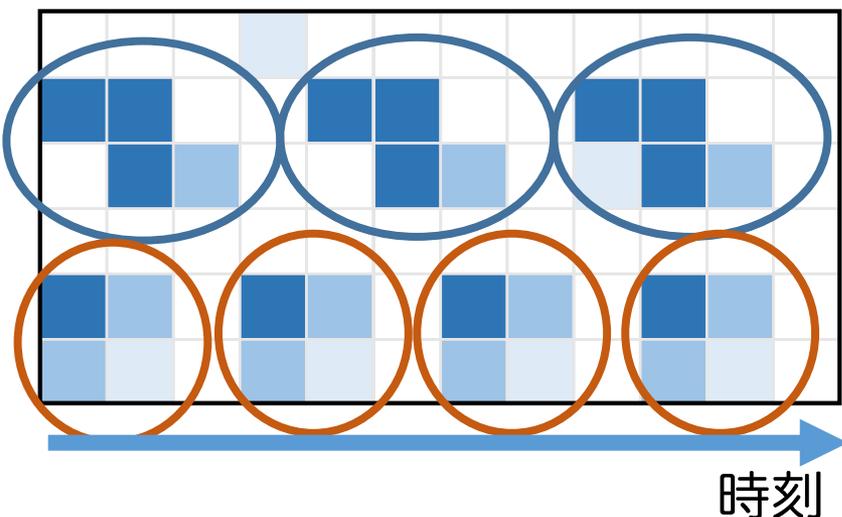
学生優秀プレゼンテーション賞対象 D-46

○鈴木 惇¹, 山西 健司^{1,2}
1: 東京大学 2: CREST

畳み込み非負値行列因子分解(NMFD)

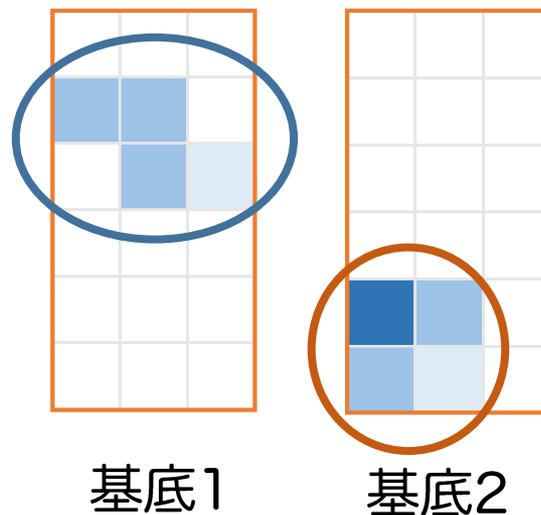
Given:

多次元時系列非負データ



Get:

時間連続な特徴を持つ非負基底



問題: 基底数・基底の時間幅の決定

難点: モデルの非正則性・各時刻の非独立性 😞

本研究: 完全変数化を伴うMDL規準を導出し難点を回避 😊

D-48 : 制約付き固有値問題を用いた

半教師付きスペクトラルクラスタリングによる画像領域分割

上田隼也(筑波大学大学院システム情報工学研究科(博士前期課程)コンピュータサイエンス専攻),

櫻井鉄也(筑波大学システム情報系), 保國恵一(筑波大学システム情報系)

画像領域分割結果 (2分割)

スペクトラルクラスタリング :

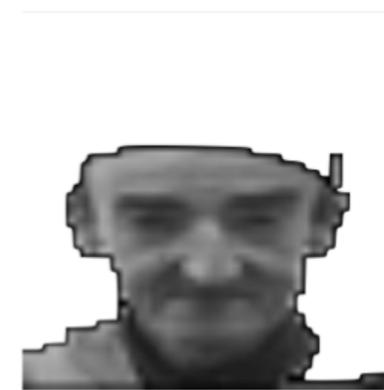
グラフ分割問題から固有値問題へ帰着
対象とするデータを固有ベクトル空間に
射影してクラスタリング

教師なし学習

原画像



領域1



領域2



提案手法 :

固有ベクトルの要素に事前知識に基づく
制約条件を課した固有値問題を解く

半教師付き学習

制約条件

中央部分が同一クラス

領域1



領域2



画像領域分割へ応用

D-50: 差分プライバシーを保証した外れ値分析アルゴリズムの高速化

発表者: 岡田莉奈・福地一斗(筑波大)・佐久間淳(筑波大/JST CREST)

学生優秀
プレゼンテーション
賞対象

データベース内の外れ値の個数のクエリに対して、差分プライバシーを保証するための評価式 [1]

$$LS_{qcount}^{(t)} \leq \max_{\mathbf{x} \in \mathcal{D}} \{ \max\{C^{(t)}(X, \mathbf{x}, k, r), C^{(t)}(X, \mathbf{x}, k-1, r)\} + t + 1 \}$$

既存アルゴリズム

全レコードの組合せを考慮する必要がある。

計算量: $\mathcal{O}(2^N)$

N は、レコード数

$N = 100, d = 2$
のとき、3週間以上
経っても計算が
終わらない!

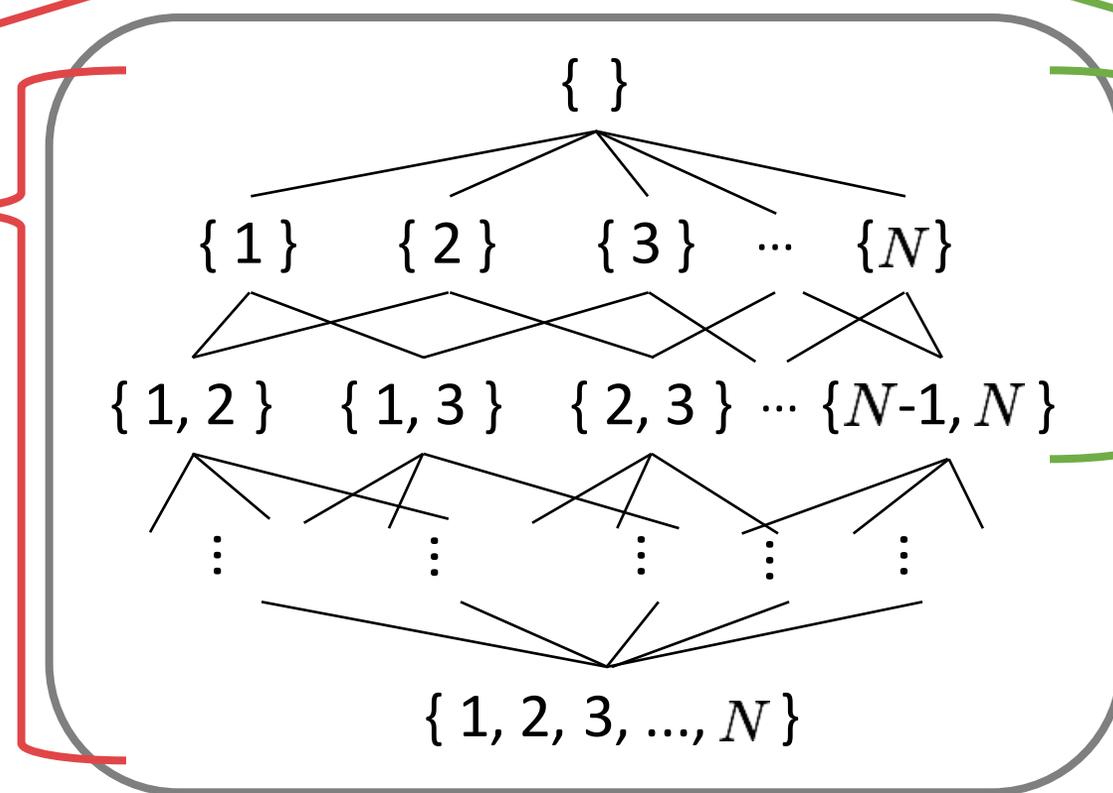
提案アルゴリズム

d 個までのレコードの組合せの考慮で済む。

計算量: $\mathcal{O}(N^d)$

d は、属性数

$N = 100, d = 2$
のとき、0.25秒で
計算が終わる!

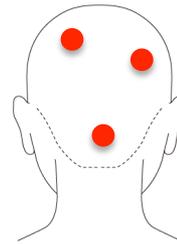
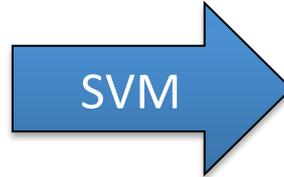
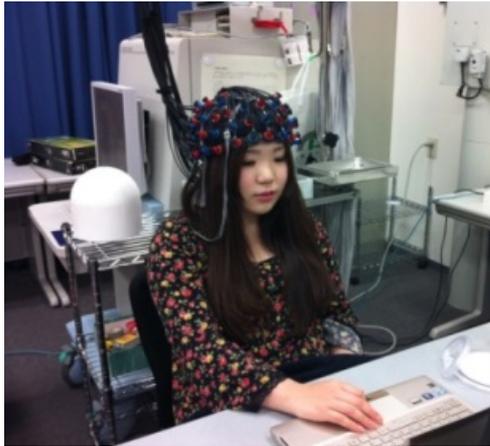


fNIRSデータからの脳の機能的結合の グレンジャー因果解析

学生優秀
プレゼンテーション対象

安部齊志(筑波大学), 佐久間淳(筑波大学/JST CREST),
滝謙一(同志社大学), 廣安知之(同志社大学)

- fNIRS:脳血流量の相対的変化を多点で計測



関連研究[Hiroyasu15]

- ✓ 活動部位の特定
- ✓ 点推定



提案

- ✓ 機能的結合の特定
- ✓ グラフ推定

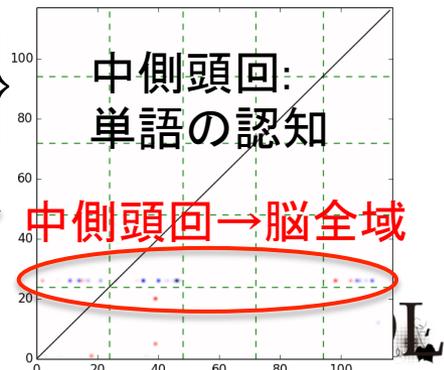
- 目的

– 被験者や正答率の違いによる機能的結合の違いを発見

$$\arg \min_{\mathbf{W}_1, \dots, \mathbf{W}_S} \left\{ \sum_{k=1}^K \sum_{s=1}^S \left\| \mathbf{Y}_s^k - \mathbf{X}_s^k \mathbf{W}_s \right\|_F^2 + \phi(\mathbf{W}_1, \dots, \mathbf{W}_S) \right\}$$

正則化検討のための予備実験

脳全体に影響を及ぼすような原因部位が存在.



Large Margin仮定における多重仮説 検定の差分プライバシー

柿崎和也(筑波大学), 佐久間淳(筑波大学/JST CREST)

属性の組み合わせを要因とする多重仮説検定において、検定統計量を元に個人の機微な情報が推測される可能性がある

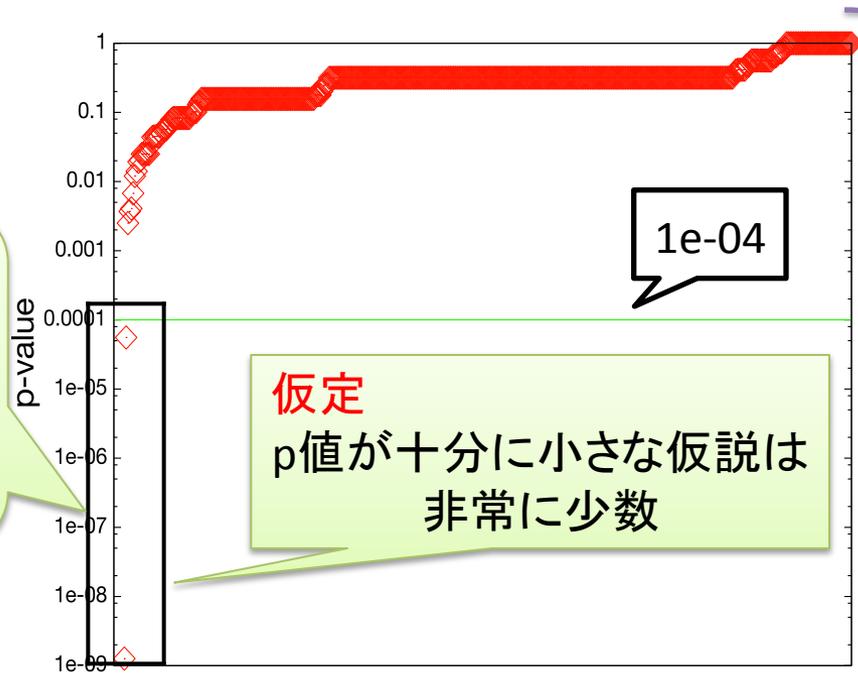
- 差分プライバシー(DP)を保証することで個人のプライバシー保護が行える

問題：仮説数が多い場合、DPを保証しつつ、検定結果の有用性を保つ検定統計量を公開することは難しい

提案：仮説に仮定を設け、少数のsignificantな検定統計量だけを抽出し公開

提案

DPを保証しながら
少数のsignificantな検
定統計量だけを抽出

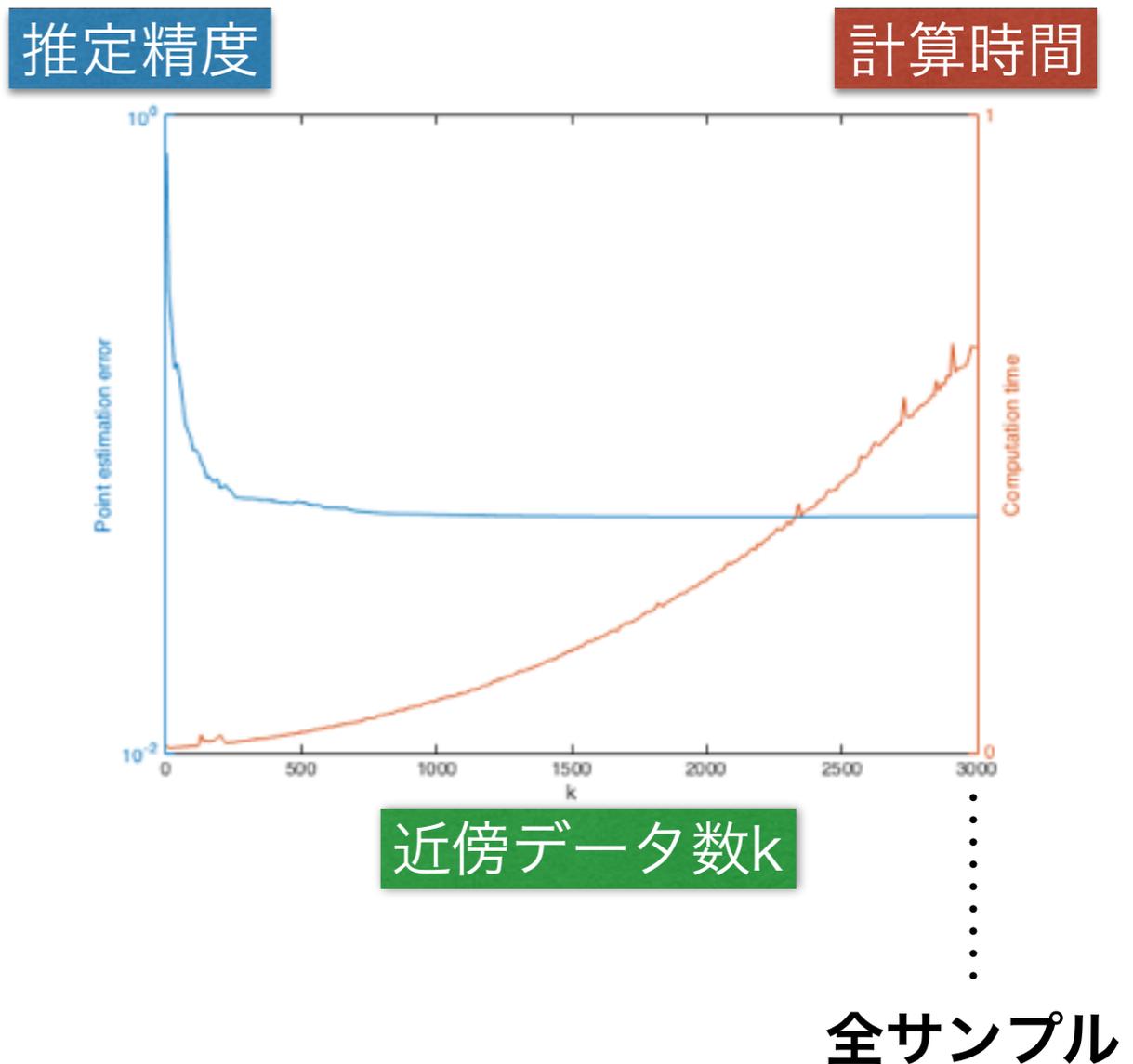
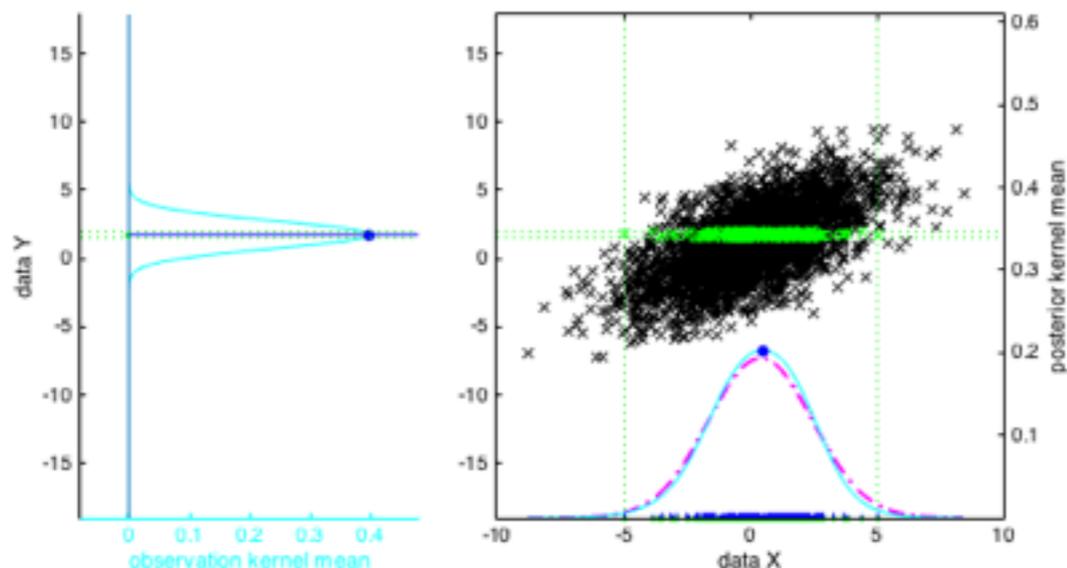


DPを保証しながら
全ての検定統計量
を公開することは難
しい

D-56: kNNを用いたカーネルベイズの 計算量削減法の検討

苗村智行 都築俊介 西山悠 (電気通信大学 情報システム学研究所)

- カーネルベイズの問題点:
サンプル数×サンプル数の
逆行列計算や行列の積を行う
- Kernel ABCに対してkNNを
組み合わせた計算量削減法を
提案する



疾患リスク公開による 遺伝情報の推定リスクの評価

草野光亮（筑波大学）・竹内一郎（名古屋工業大学）
・山田芳司（三重大学）・佐久間淳（筑波大学 / JST CREST）

背景情報

性別
体重
喫煙歴
etc...

x

遺伝情報



h

背景

10疾患の疾患リスクが公開される

疑問

遺伝情報はどの程度推定可能か？

手法

攻撃者の視点で実際に推定を行う
→ 列挙やMCMC

目的

遺伝情報の漏れにくい丸め関数の検討

$$f_{\text{疾患A}}(x, h)$$

推定
(攻撃者)

ロジスティック回帰

疾患リスク

$$y_{\text{疾患A}} = 0.878825\dots$$

丸め処理 ↓ $y_{\text{疾患A}}^{(L)} \leq y_{\text{疾患A}} \leq y_{\text{疾患A}}^{(U)}$

公開疾患リスク

$$\left(y_{\text{疾患A}}^{(L)}, y_{\text{疾患A}}^{(U)} \right)$$

ex) 「あなたの疾患リスクは A です」 or グラフ

D-60 クラウドソーシングワーカーの 自己申告属性を用いた回答統合法

馬場 雪乃（京大）、清水 伸幸、藤田 澄男（Yahoo! JAPAN研究所）

SNSアカウントの属性推定タスクにおける回答統合法を提案
ワーカーの自己申告属性を利用、嘘つきワーカーの存在を考慮



Q.この人の性別は?

A.女性 (Pumpkin icon) A.女性 (Cat icon) A.男性 (Ghost icon)

Q.この人の最終学歴は?

A.大学 (Pumpkin icon) A.高校 (Cat icon) A.大学院 (Ghost icon)

Q. あなたの属性は?
A. 男性、40代、既婚、院卒

Q. あなたの属性は?
A. 女性、30代、未婚、大卒

Q. あなたの属性は?
A. 男性、50代、既婚、高卒

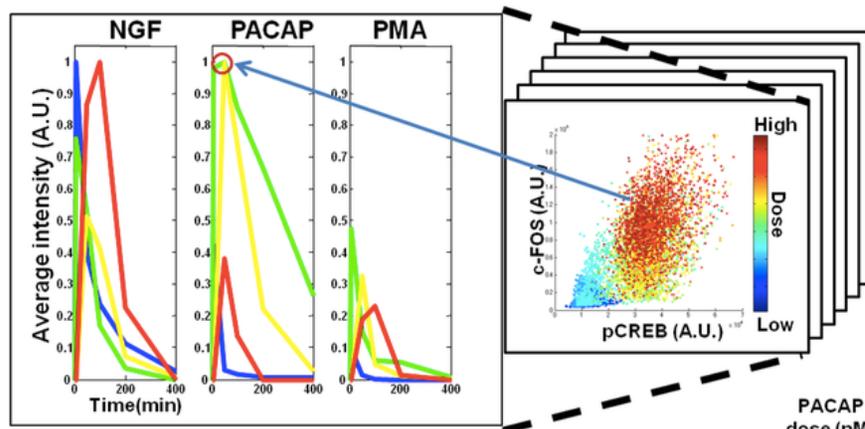
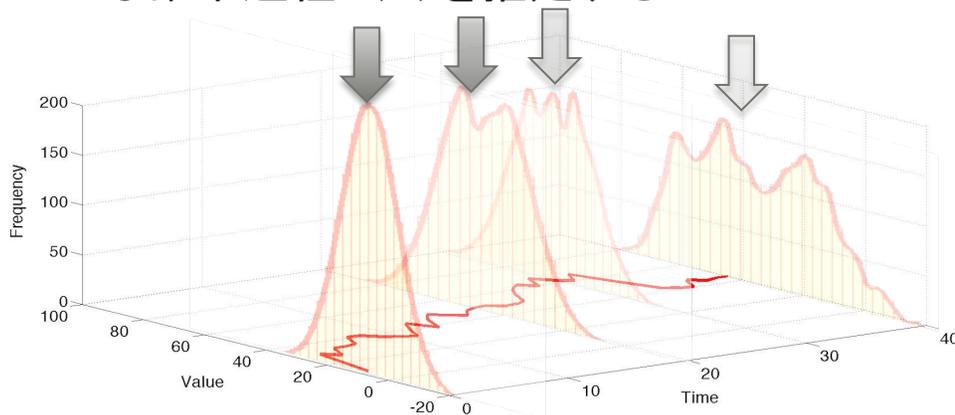
本研究はJST, ERATO, 河原林巨大グラフプロジェクトの支援を受けたものです

D-62

時系列の周辺分布からの EM型モデル推定

小山雅典 前田新一(京大)

軌道が失われていて、周辺分布 $p(x(t))$ のデータのみがある状態で、パラメトリックな確率過程 $X(\theta)$ を推定する



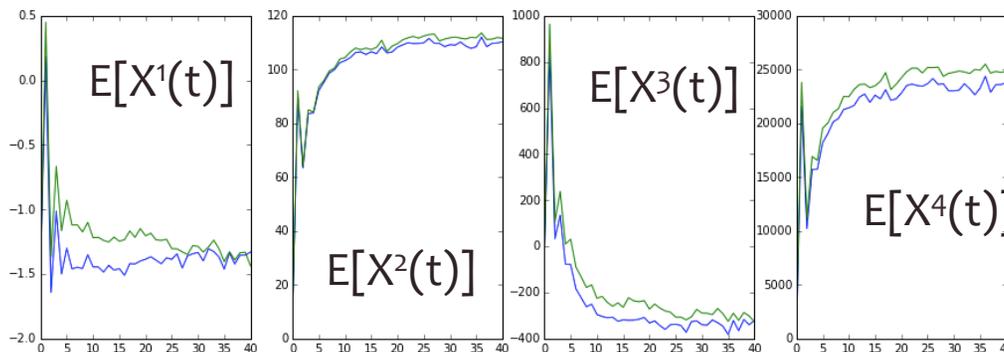
提案:

- (1)すべての粒子がすべての観測点に対応するパーティクルフィルタ
- (2)分布の分布であるディルクレ分布の事後分布の近似

を使ったEM法

詳細はポスターにて

時系列の特定に対してどのモーメントが重要か、自動的にわかる



タンパク質構造予測のための表現学習と深層学習

タンパク質大規模データベース

ENIKLGFGGLGQGSALAHGI
 ANANIKKENLFYYG
 NSEIDNLLKDKYGEQLGV
 EFFNQIN
 ADQRAAEQYVIDEYNKLLK
 TRLREAALA
 ...

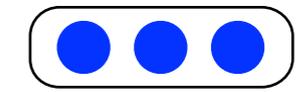
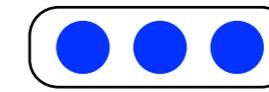
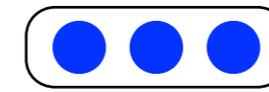
+ **表現学習** →

N-gramアミノ酸ベクトル

GFG

DKY

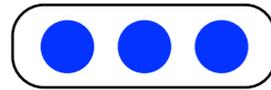
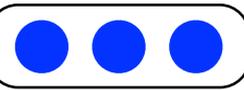
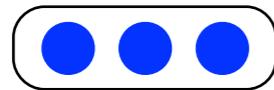
ADQ



- 低次元の**タンパク質ベクトル**を構成
- 構造クラスを**高精度かつ高速**に分類
- **教師なし**で構造の学習が可能

タンパク質の**立体構造**を学習

深層学習

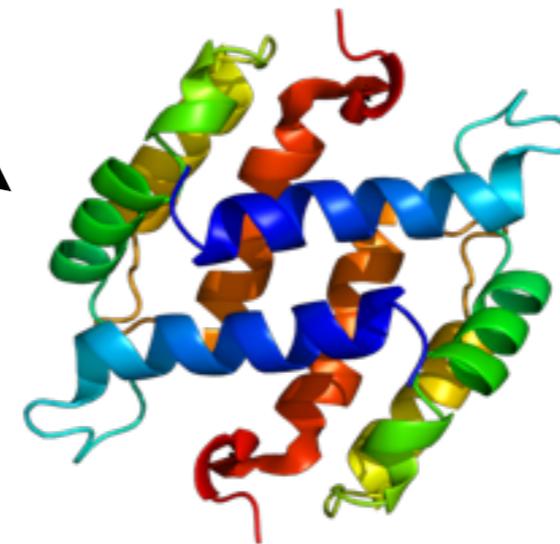


GFG

DKY

ADQ

アウトプットは各アミノ酸の
 三次元空間における位置関係



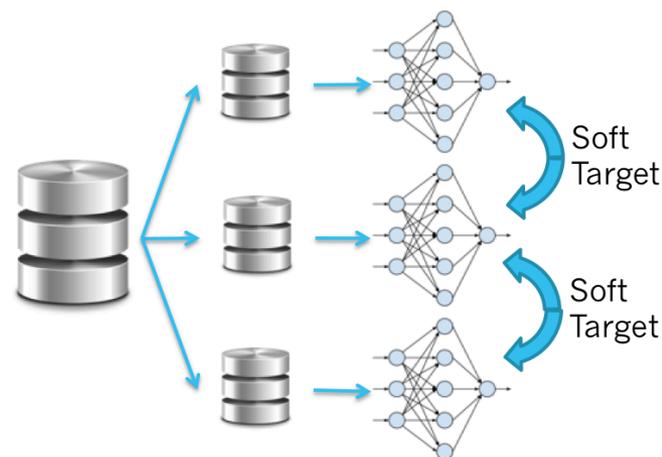
- 低次から**高次**の立体構造を学習し予測
- 現在の最新技術でも**20%程度**の精度
- 教師あり深層学習で上記のベクトルを**再学習**

D-66 Community Learning : 蒸留を用いた分散深層学習と化合物活性予測マルチタスク学習への応用

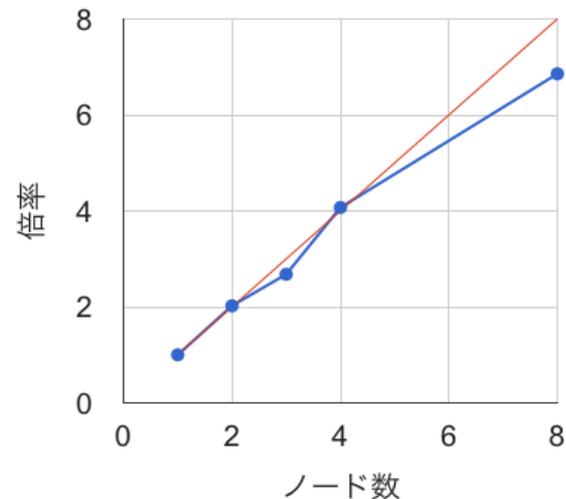
大野健太、岡野原大輔(Preferred Networks)、大田信行 (Preferred Networks America)

- 分散深層学習の新手法を提案

	既存手法	提案手法
ノード間の通信データ	ニューラルネット(NN)のパラメータ	NNの分類結果 (Soft Target)
ノード間通信量	× ノード間通信がボトルネック	✓ 通信量の削減
NN構造への制限	× 全ニューラルネットが同一構造	✓ 入出力の形式が一致すれば良い



- 化合物活性予測のタスクに適用。8ノード・24GPUでのスケーラビリティ・AUC向上を確認

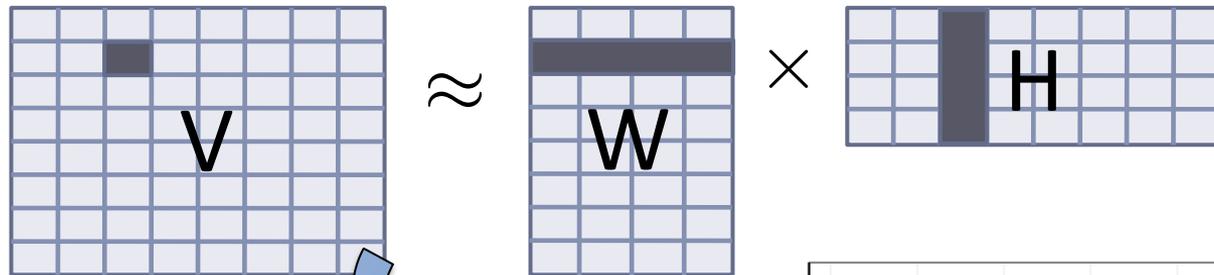


D-68 切断正規分布に基づく非負値行列分解

白川真一(筑波大学), 鈴木創, 大原剛三(青山学院大学)

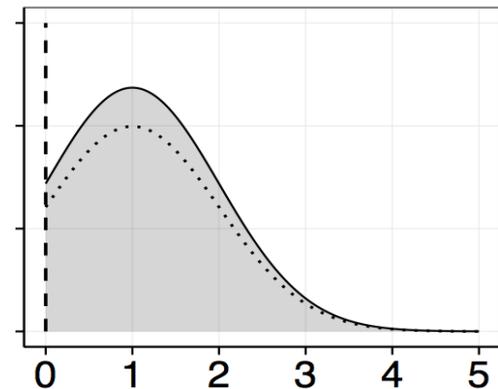
非負値行列分解(NMF)

- 行列要素の非負値制約もと分解
- 二乗誤差NMF = 正規分布仮定(負値領域にも確率密度を仮定)



切断正規分布

- 確率変数の定義域が有界
- NMFにより適しているのでは



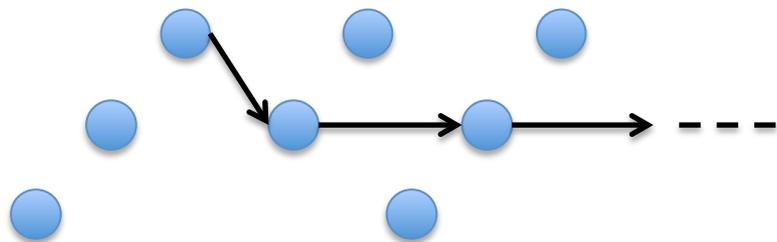
➡ 切断正規分布に基づくNMFの学習法の導出と評価

D-70

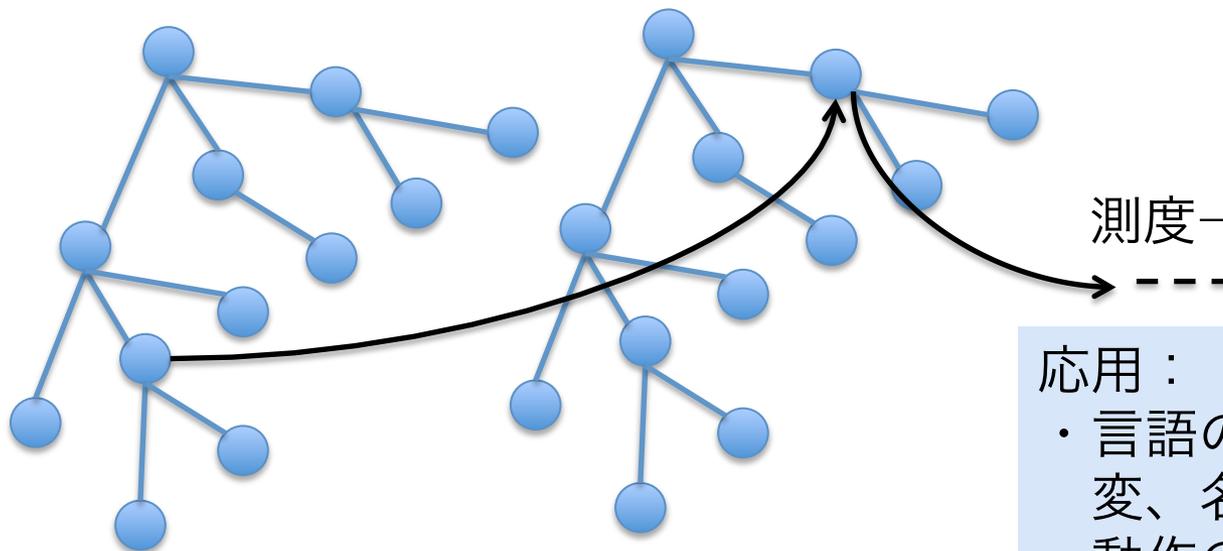
The Infinite Tree Hidden Markov Model

持橋大地 (統数研)

HMM (infinite HMM)



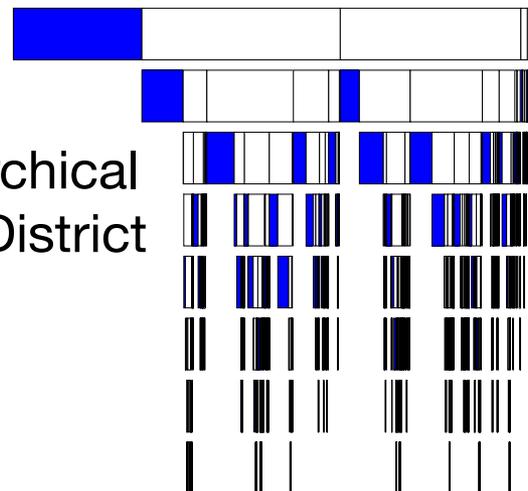
Infinite Tree HMM



- HMMの本質的な拡張
- 深さ1に制限すれば, infinite HMMと等価

Tree-structured stick-breaking process (Adams+ NIPS 2010)

+階層化
(via hierarchical
Chinese District
Process)



測度→測度への精密な写像

応用：

- 言語の階層 (動詞-他動詞-サ変、名詞-一般名詞-人名)
- 動作の階層 (歩く-速く-滑らか)
- その他色々な応用が可能

D-72 統計調査の約600分類の符号付与システムについて

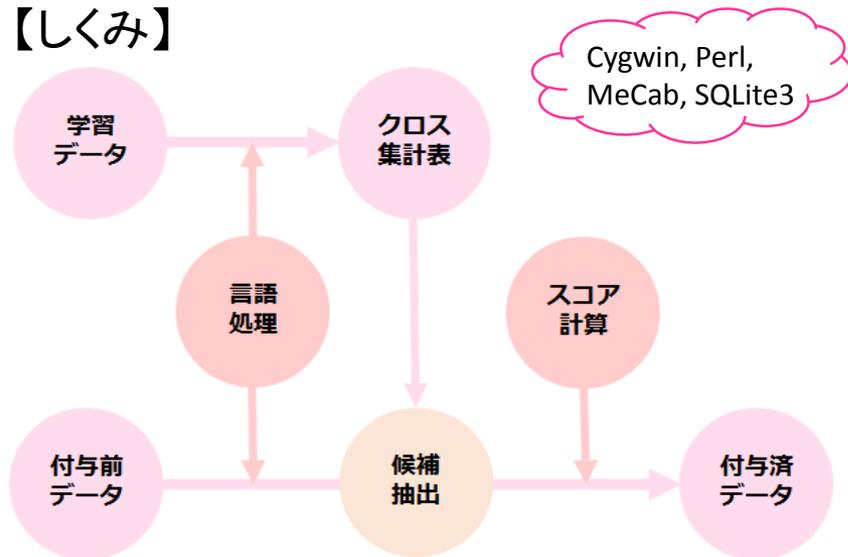
床 裕佳子, 下野 寿之, 和田 かず美, 坂下 佳一郎 (独立行政法人 統計センター)

2015/11/26 IBIS2015(つくば国際会議場)

【目的】

単純な仕組みで高精度かつ処理速度の速い
符号付与システムを作る。

【しくみ】

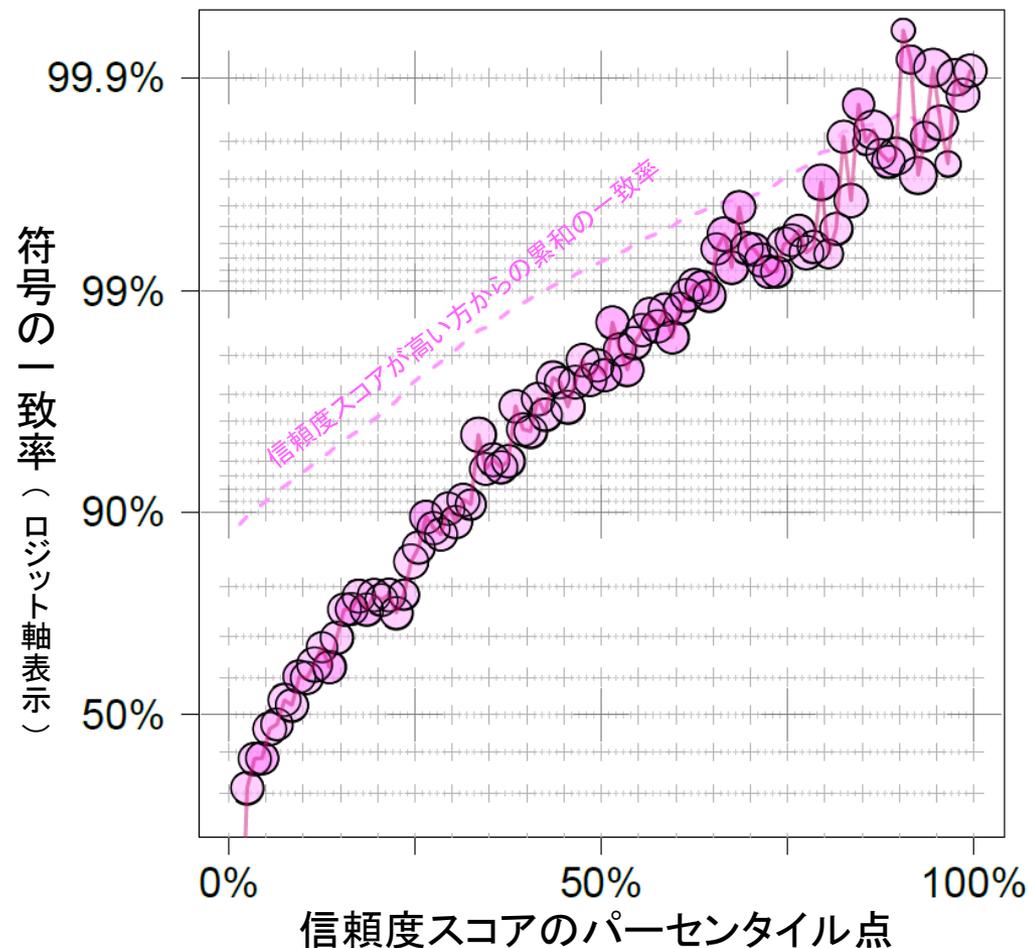


約187万件で学習して、約32万件を評価した。

処理時間 1万件の学習に約2秒
(Xeon, 3GHz) 1万件の評価に約5秒

【結果】

信頼度スコアのパーセンタイルごとの一致率



まだ不安な点

- ・複数の意味を持つ単語 / “強い”キーワードの副作用
- ・表記ゆれ / カタカナ連続記入 など

介護レセプトデータに対する横断的特徴選択による介護サービスの評価

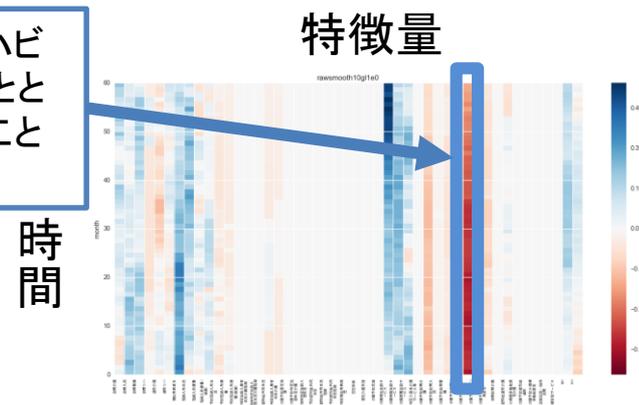
中里佳央(筑波大学),佐久間淳(筑波大学/JST CREST),川村颯(筑波大学医学医療系ヘルスサービスリサーチ分野),田宮菜奈子(筑波大学医学医療系ヘルスサービスリサーチ分野)

- 介護レセプト: 介護サービス利用量と要介護度の時系列データ
- 目的: 特徴選択により要介護度の推移と介護サービスの効果の関連性を発見

$$J = \arg \min_{B_t \in T} \sum_{t=1}^T \| S \cdot (B_t X_t - Y_t) \|_F^2 + \sum_{i=1}^c \sum_{t=0}^{T-1} \| B_{t+1,i} - B_{t,i} \|^2 + \sum_{d=1}^D \sqrt{\sum_{t=1}^T \| B_{t,d} \|^2}$$

- 欠損値への対応
- 時間平滑性への対応

介護予防通所リハビリを受けていることと要介護度が低いことに大きな関連性



マルチタスク学習による時間横断的な介護サービスの評価

混合ガウシアングラフィカルモデルの学習手法と交通状態補間

鈴木惇平・原祐輔・桑原雅夫(東北大学大学院情報科学研究科)

学生優秀プレゼンテーション賞対象

◆ 研究目的

- 道路ネットワーク上で部分的に観測されたリンク旅行速度データによる未観測リンク速度の補間し, ネットワーク全体の交通状態を把握する

◆ 研究概要

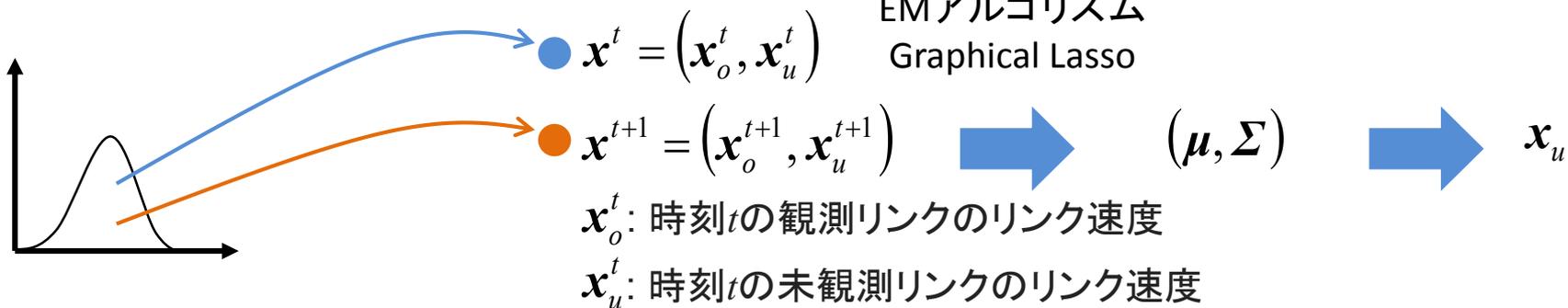
観測データの生成過程

観測データ

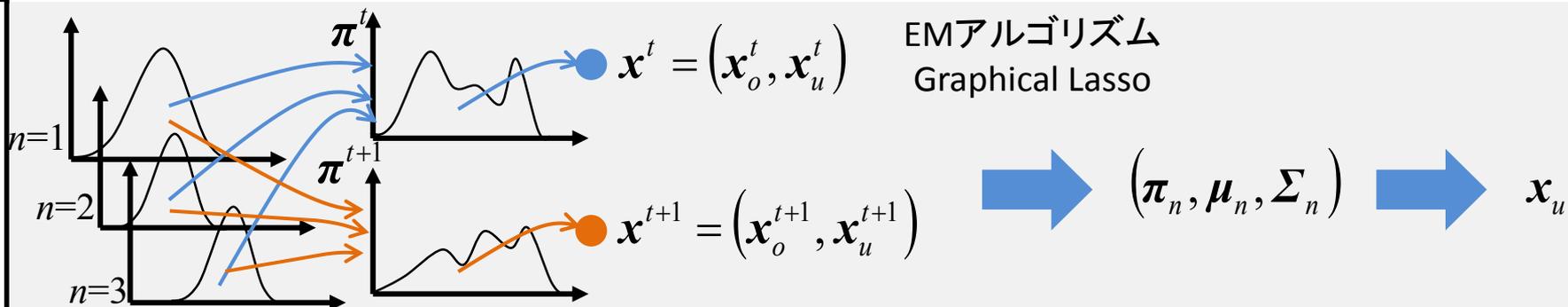
学習

補間

既往研究



本研究



D-84: Barron and Cover 理論の教師付き学習への拡張及び lasso への応用

川喜田雅則 (九大), 竹内純一 (九大)

Barron and Cover 理論 Barron and Cover (1991)

- ▶ MDL 理論の数学的正当化として最も有名.
- ▶ 有限の標本数における MDL 推定量の精密なリスクの上界評価が可能.
- ▶ 確率変数の有界性やモーメント条件を仮定しなくて良い.
- ▶ パラメータ空間が可算空間でないといけない.

最近の進展

- ▶ いつの間にかパラメータ空間が連続でも適用可能にする工夫が提案されていた! (Barron et. al. (2008)) また, 実際にその工夫を適用して lasso のリスク上界が導出された.
- ▶ しかし実はこの工夫は固定計画が前提. 乱数計画に拡張することは本質的に困難.

主要な結果

- ▶ 共変量が正規分布のときに, この問題を解決し lasso の汎化誤差の精密な上界を導出した. 我々の方法は教師付き学習に BC 理論を適用する際の標準的な方法の一つになると期待される.

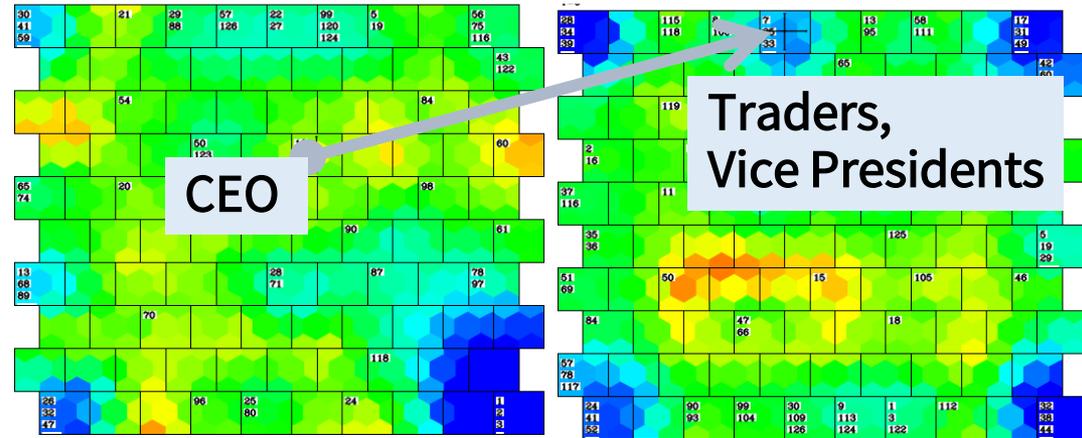
Tensor SOMによる E-mailデータのTopic・Roleの同時可視化

本手法では、

- BoWデータからメッセージ、単語の潜在変数を推定する
- メッセージの潜在変数の分布から、送信者と受信者の潜在変数を推定する

例えば、

指定した送信者と受信者の
トピック・ロールを知りたい



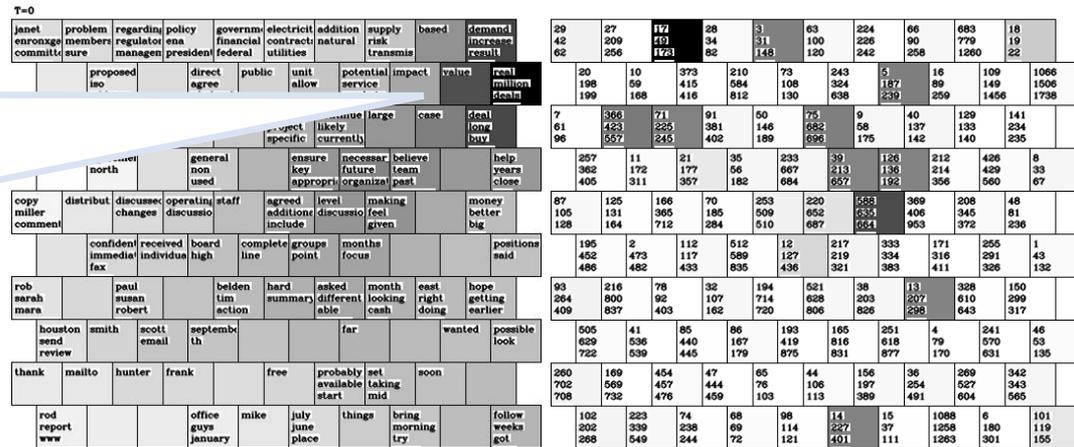
送信者マップ

受信者マップ

risk, cost, trading, ...

million, prices, values, buy, ...

future, decision, making, ...



単語マップ

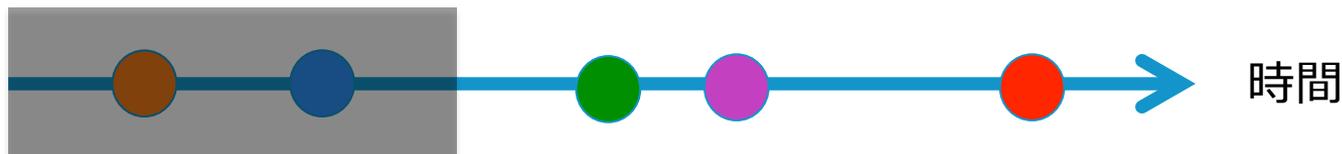
メッセージマップ

Web上の消費行動からのオフラインイベントモデリング

小山田創哲（リクルートテクノロジーズ） 小山雅典（京都大学）

問題

部分的に観測されたイベントの系列から、真のイベント系列の分布および、イベント間の関係性を推定する



- 設定:
1. 各イベントは一度しか発生しない
 2. あるイベントの発生は別のイベントの発生確率へ影響を与える
 3. 観測が部分的に（途中から）与えられる

アプローチ

連続時間マルコフ過程を用いてモデリングし、EMアルゴリズムで推定を行った

適格度の履歴を用いた Natural Actor-Critic 法の解析と 頑健な学習法の提案

岩城諒 (大阪大学)
横山裕樹 (大阪大学)
浅田稔 (大阪大学)

逐次的な自然勾配推定の学習則 [木村 2007] を解析

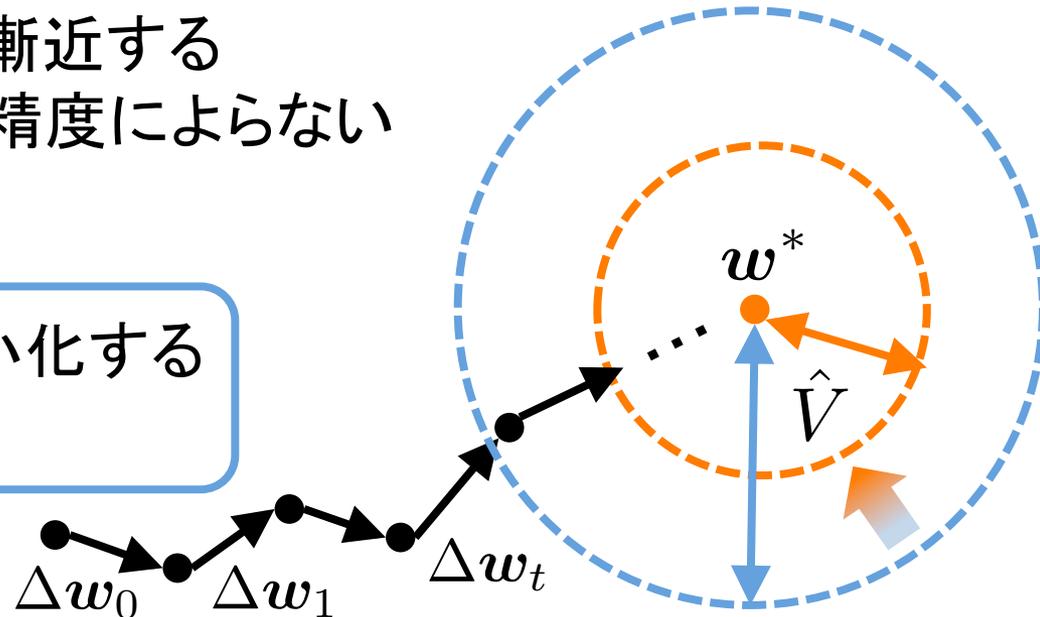
$$E \left[\sum_{t=0}^{\infty} \Delta w_t \right] \approx \gamma \sum_{t=0}^{\infty} \nabla_{\theta} E \left[\sum_{t=0}^{\infty} \gamma^t r_t \mid \theta \right]$$

逐次更新の期待値

- ✓ 収益の期待値の勾配に漸近する
- ✓ Critic による価値推定の精度によらない

更新の分散を明示的に最小化する
Critic の学習則の提案

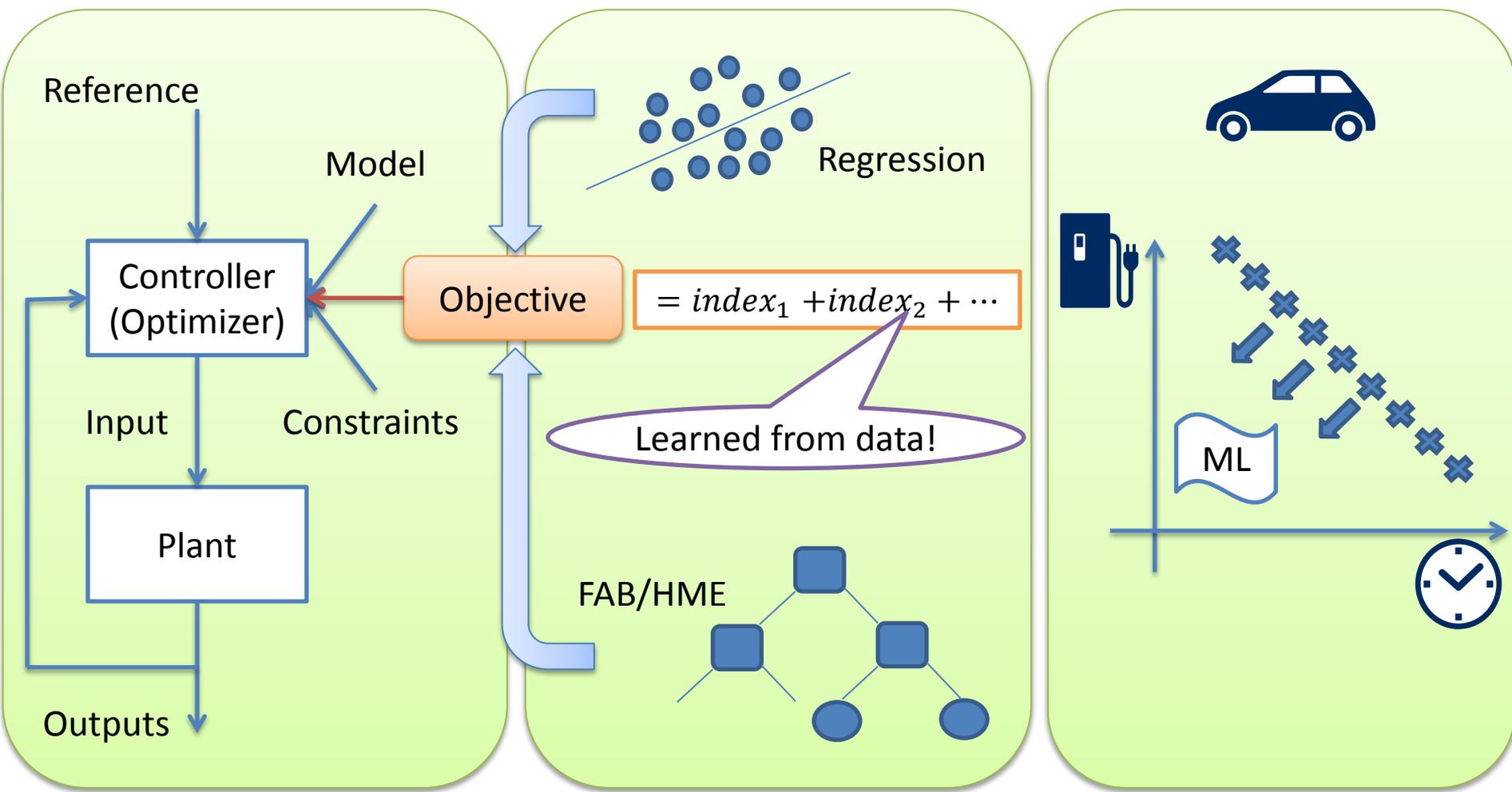
学生優秀プレゼンテーション賞対象



Predictive Control Problem Design Using Machine Learning Techniques

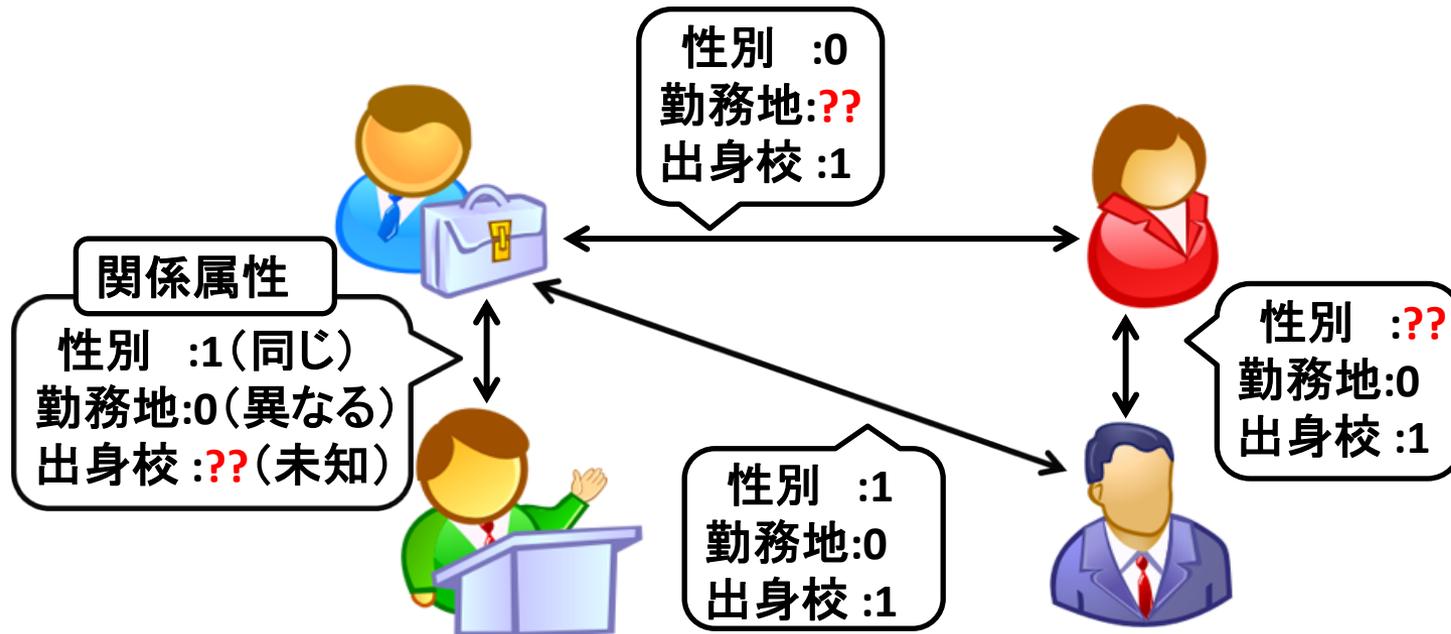
Wemer M. Wee, Riki Eto and Yoshio Kameda
NEC Corporation

Model Predictive Control + **Machine Learning** → **Autonomous Driving**



【D-94】双対分解を用いた無限潜在特徴モデルによる 関係データの属性予測

西出飛翔(神戸大), 江口浩二(神戸大)



目的

- SNSなどの複数の属性が与えられた関係データから、既知の属性ラベルを利用しつつ潜在特徴を推定し、未知の属性の予測を行う

提案手法

- 複雑な目的関数の最適解を求める双対分解をMedLFRM[Jun Zhu, 2012]に適用
- 複数の属性に対するマルチタスクラーニングによってモデルを推定

実験

- 双対分解の適用による予測精度の向上を評価

[D-96] 視線運動からの商品カタログ閲覧状態の推定

発表：石川恵理奈（松山研究室，京都大学大学院情報学研究科）

学生優秀プレゼンテーション賞対象

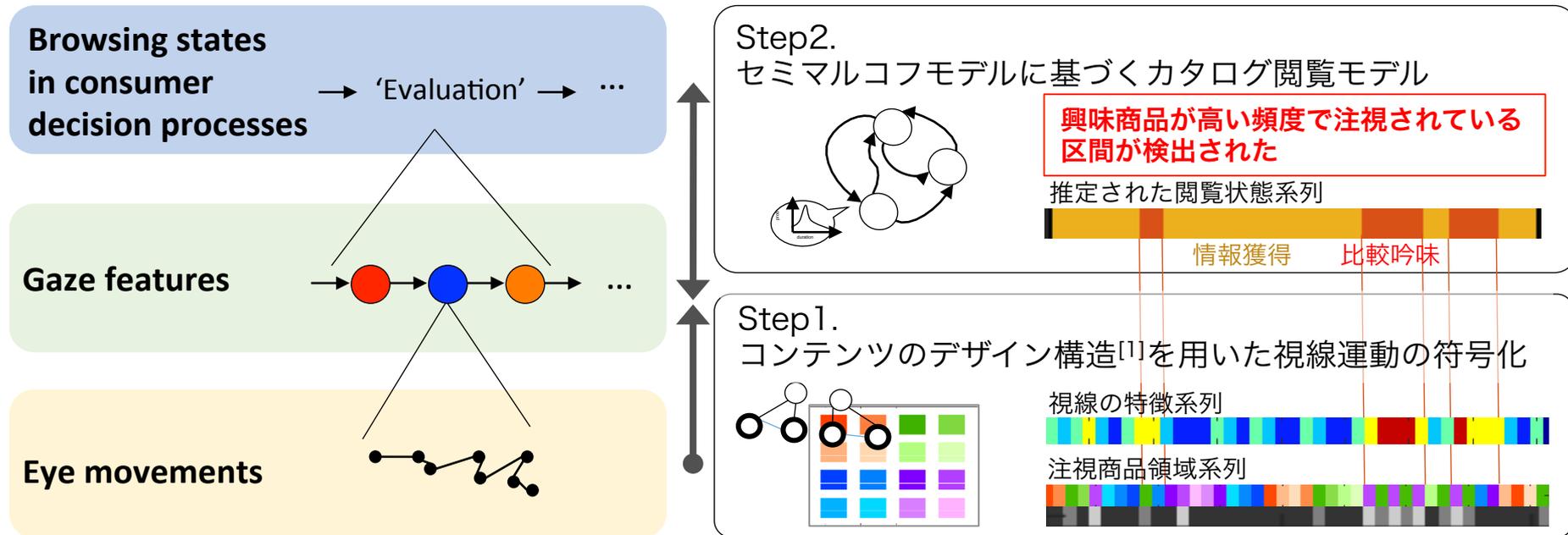
「目は心の窓」 デジタルカタログ閲覧時の意思決定過程を閲覧者の視線から推定する

ポイント

注視の「多義性」の解消

(1) 局所的な文脈情報：注視対象間の意味的・空間的情報の利用

(2) 大域的な文脈情報：商品情報の獲得・処理過程の推定・利用



様々な応用先・関連分野：インタラクティブ情報システム，消費者行動解析，興味推定，etc.

関連文献

- [1] Erina Ishikawa, Hiroaki Kawashima, and Takashi Matsuyama: Using Designed Structure of Visual Content to Understand Content-Browsing Behavior, IEICE Transactions, 98(8), 2015.
- [2] Erina Ishikawa Schaffer, Hiroaki Kawashima, and Takashi Matsuyama: Estimation of Browsing States in Consumer Decision Processes from Eye Movements, The 3rd IAPR Asian Conference on Pattern Recognition (ACPR2015).

D-83 Latent Synonym Topic Models

Weihua Hu(UTokyo), Motoki Sato(NAIST), Hideki Mima(UTokyo), Jun'ichi Tsujii(AIST)

	LDA	Latent synonym Topic Models
Topic as a distribution over	words.	latent synonyms.
Topic inference depends on	co-occurrence of words.	co-occurrence of latent synonyms.
Vocabulary size is	large and sparse. (Usually 5000 - 10000)	small and dense. (Usually 500 - 1000)

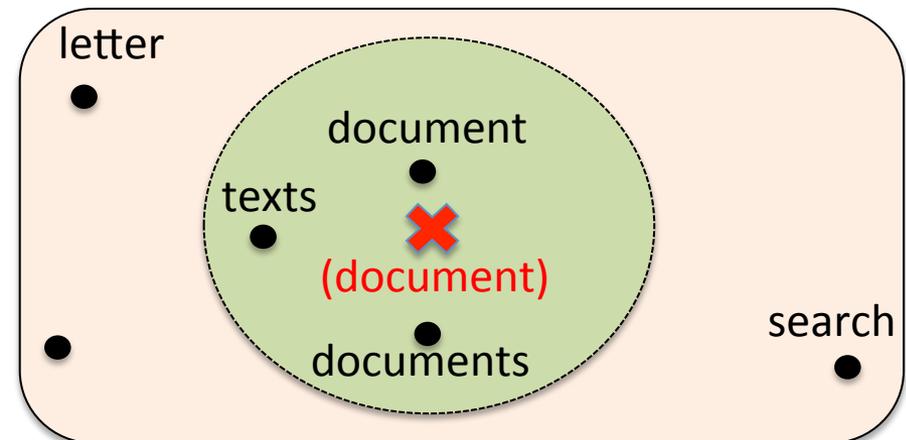
Observation

Synonyms are clustered in **word2vec** vector space.

Proposal

We introduce **latent synonym vectors** from which actual word vectors are generated with some noise.

In word vector space ...



✖ : synonym vector ● : word vector