D1-1

動的モード分解のための Koopman不変空間学習

武石 直也（東大）
河原 吉伸（阪大，理研）
矢入 健久（東大）

非線形力学系　　無限次元線形力学系

🔍 $\mathcal{K}$のスペクトルによる力学系の分解 ＝ 動的モード分解

→ **力学系解析・制御，時系列データ分析**などに利用

✏ ただし，非線形力学系ではデータの非線形変換が必要

→ 本研究では**必要な非線形変換をデータから学習する**

D1-1

動的モード分解のための Koopman不変空間学習

武石 直也（東大）
河原 吉伸（阪大，理研）
矢入 健久（東大）

非線形力学系　　無限次元線形力学系

🔍 $\mathcal{K}$のスペクトルによる力学系の分解 ＝ 動的モード分解

→ **力学系解析・制御，時系列データ分析**などに利用

✏ ただし，非線形力学系ではデータの非線形変換が必要

→ 本研究では**必要な非線形変換をデータから学習する**

D1-2: 分布に対する位相的データ解析

草野元紀(東北大・理)　【学生優秀プレゼンテーション賞対象】

位相的データ解析：データの“形”をパーシステントホモロジーという特徴量に変換する。

Q. 液体の原子配置とガラスの原子配置を区別できる？(ランダムに見える点群データの分類)

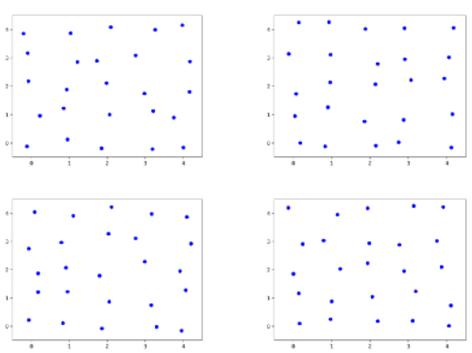
A. パーシステントホモロジーのカーネル法により区別した。 [IBIS2016]

↓

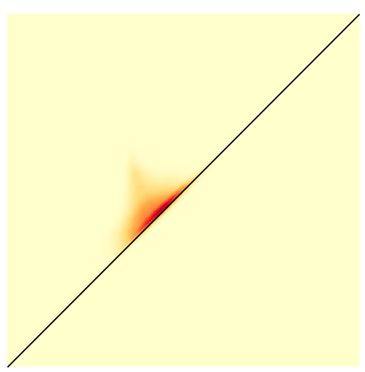
Q. 原子配置は同じなんだけど、異なる分布で揺れている。(熱揺らぎを想定している)

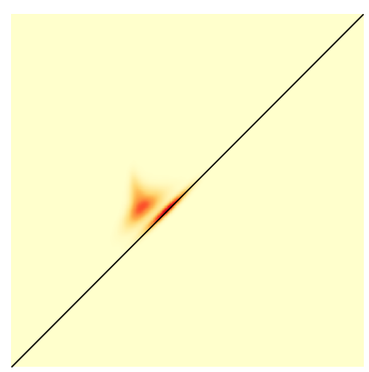
原子配置の形の情報は保持しつつ、分布の違いを評価できる？

A. **分布に対するパーシステントホモロジー**を計算することで、違いを評価できる[IBIS2017]

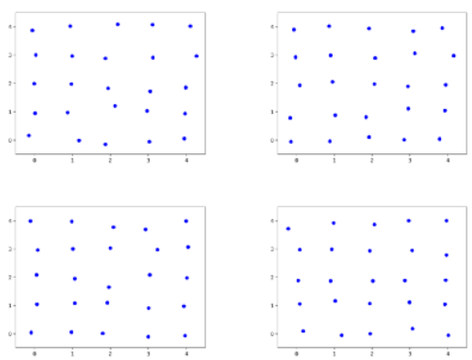


正方格子を中心とした
一様分布





分布に対するパーシステントホモロジー



正方格子を中心とした
正規分布

D1-2: 分布に対する位相的データ解析

草野元紀(東北大・理)　【学生優秀プレゼンテーション賞対象】

位相的データ解析：データの“形”をパーシステントホモロジーという特徴量に変換する。

Q. 液体の原子配置とガラスの原子配置を区別できる？(ランダムに見える点群データの分類)

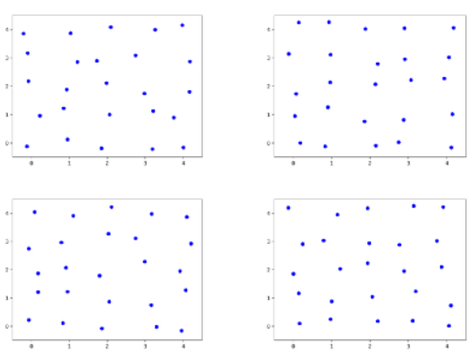
A. パーシステントホモロジーのカーネル法により区別した。 [IBIS2016]

↓

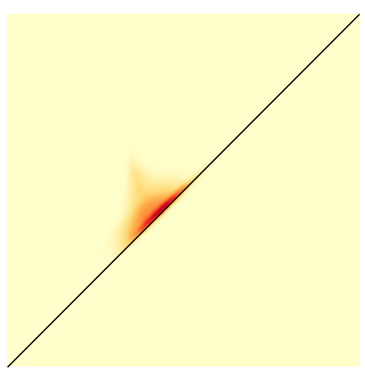
Q. 原子配置は同じなんだけど、異なる分布で揺れている。(熱揺らぎを想定している)

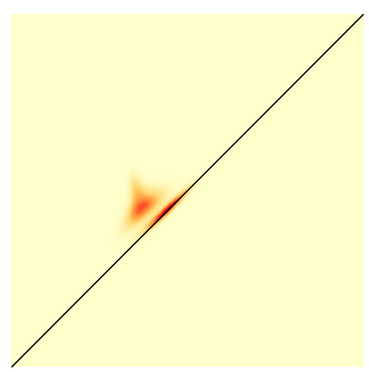
原子配置の形の情報は保持しつつ、分布の違いを評価できる？

A. **分布に対するパーシステントホモロジー**を計算することで、違いを評価できる[IBIS2017]

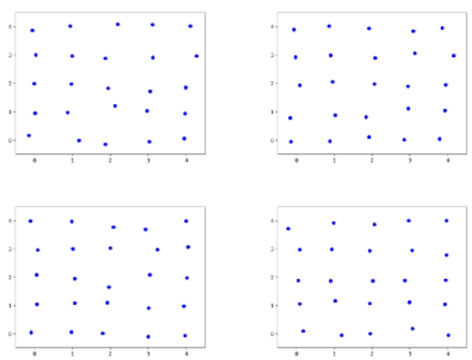


正方格子を中心とした
一様分布





分布に対するパーシステントホモロジー



正方格子を中心とした
正規分布

D1-4

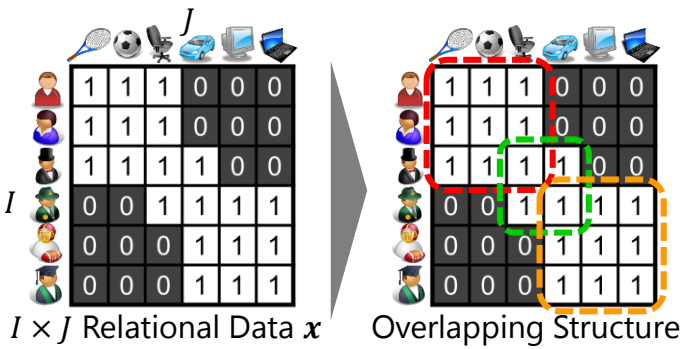
Gamma Process Edge Partition Model のモデル縮退効果の改良

To appear in NIPS2017

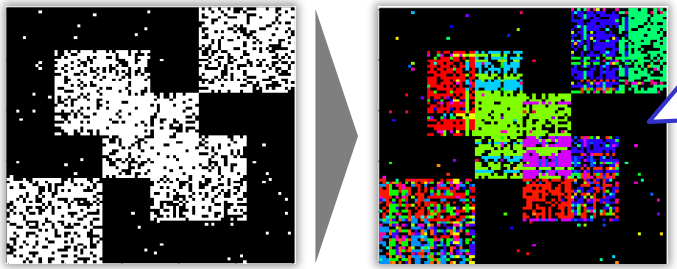
大濱 郁 (パナソニック/北大), 佐藤 一誠 (東大/理研), 喜田 拓也 (北大), 有村 博紀 (北大)

背景: 二値行列の重複構造分析

- **ΓP Edge Partition Model (EPM)** [Zhou, 15]
  - 二値行列に潜在する重複構造を抽出
  - 有効な基底数を自動的に推定することが可能



- **EPMの問題点**
  - モデル縮退が働かず有効基底数を過大評価
  - 結果として過学習を起こしやすい



98個もの基底が有効と判断

主結果: モデル縮退効果の解析と改良

- **EPMのIntensity Function (IF)の解析**

EPMの生成モデル

$$x_{i,j} = \underbrace{\mathbb{I}(m_{i,j,\cdot} \geq 1)}_{\text{(二値化)}}, \quad m_{i,j,\cdot} \sim \text{Poisson}\left(\sum_{k=1}^{T} U_{i,k} V_{j,k} \lambda_k\right),$$

$$\lambda_k \sim \underbrace{\text{Gamma}(\gamma_0/T, c_0)}_{(T\text{-Truncated ΓP})}, \quad U_{i,k} \sim \underbrace{\text{Gamma}(a_1, b_1)}_{\text{(gamma prior)}}, \quad V_{j,k} \sim \underbrace{\text{Gamma}(a_2, b_2)}_{\text{(gamma prior)}}.$$

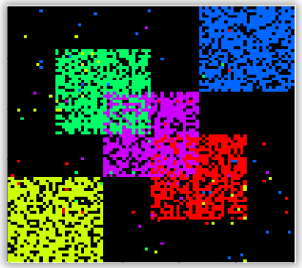
Intensity Function (IF)

$$\mathbb{E}\left[\sum_{k=1}^{\infty} U_{i,k} V_{j,k} \lambda_k\right] = \frac{a_1}{b_1} \times \frac{a_2}{b_2} \times \frac{\gamma_0}{c_0}$$

3つの非負値項の積

超パラメタの選択次第で縮退が妨害されることが判明

- **効果的にモデル縮退が働くEPMの構成を複数提案**
  - CEPM: ガンマ事前分布に制約を与えるモデル (Gibbs)
  - DEPM: 解が一意に決まるように問題を変換したモデル (Gibbs)
  - **IDEPM: DEPMの無限極限を取ったモデル (Collapsed Gibbs)**


IDEPM Solution

IDEPMが汎化性能，リンク予測性能，計算速度の全てにおいて最高性能を示すことを確認 ☺

ArXiv Paper: https://arxiv.org/abs/1709.08770

D1-4

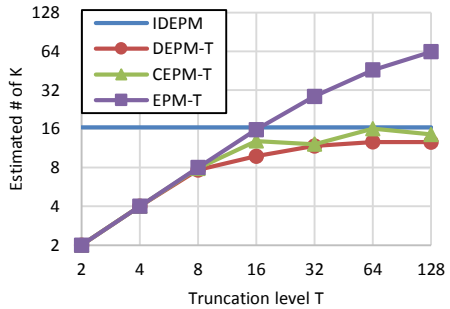
Gamma Process Edge Partition Model のモデル縮退効果の改良

To appear in NIPS2017

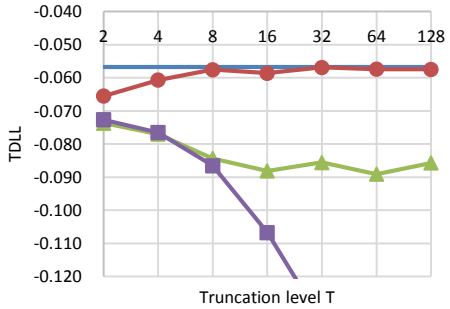
大濵 郁 (パナソニック/北大), 佐藤 一誠 (東大/理研), 喜田 拓也 (北大), 有村 博紀 (北大)

実験結果 (Enronデータの結果のみ抜粋)

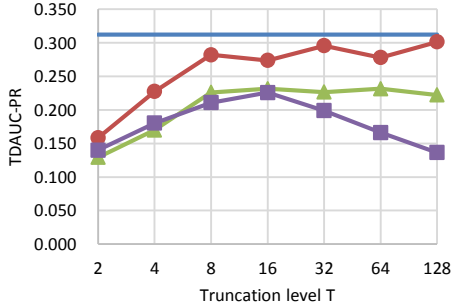
■モデル縮退効果
提案モデルであるCEPM, DEPM, IDEPMは正しくモデル縮退効果が働くことを確認☺



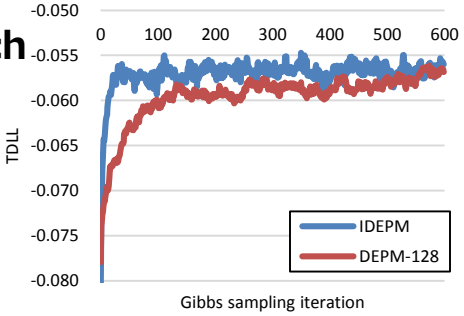
■汎化性能
提案モデルは汎化性能が改善．特にIDEPMは最も高い汎化能力を示すことを確認☺



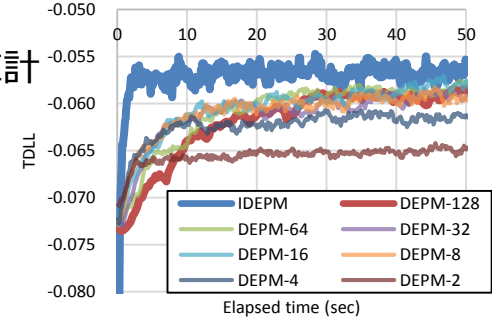
■リンク予測性能
提案モデルはリンク予測性能が改善．特にIDEPMは最も高いリンク予測能力を示すことを確認☺



■テスト尤度 vs # of Epoch
IDEPMはDEPMよりも効率的に収束することを確認☺



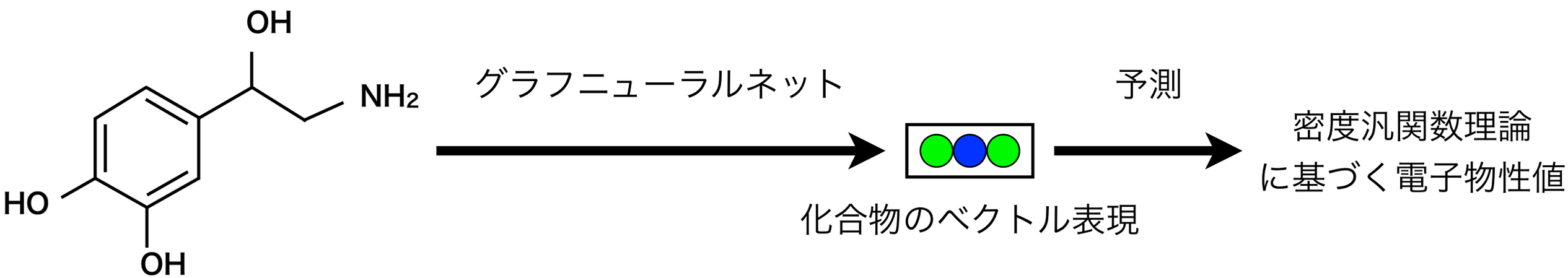
■テスト尤度 vs 経過時間
IDEPMはDEPMよりも高速に計算できることを確認☺



結論:
- EPMのモデル縮退が働かない理由を解析
- 解析で得た知見に基づきCEPM, DEPMを提案
- 近似無しの無限モデルであるIDEPMを提案
- IDEPMが汎化性能，リンク予測性能，計算速度の全てにおいて最高性能を示すことを確認

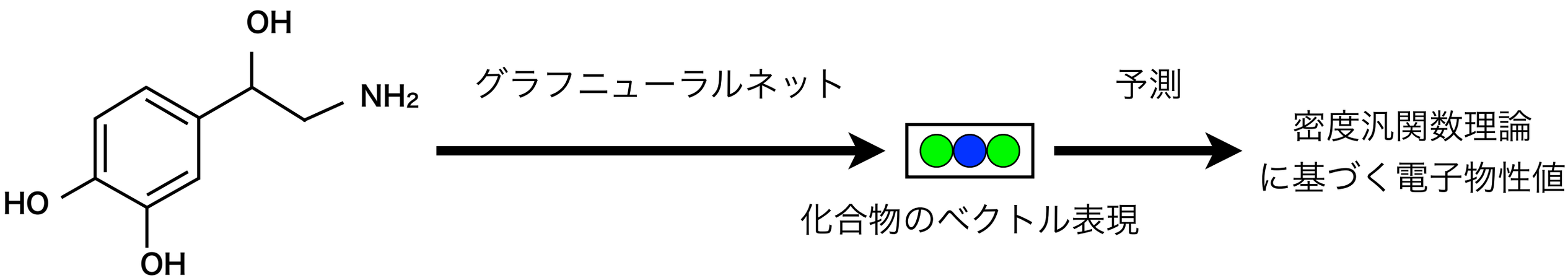
ArXiv Paper: https://arxiv.org/abs/1709.08770

機械学習を用いた量子化学計算の高速かつ高精度な近似法



- **化合物の何らかの性質を反映する**低次元かつ連続値のベクトル表現を獲得
- 化学・物理学に関する知識・特徴量を一切用いない
  データ駆動型グラフニューラルネットワークの一貫学習を提案
- 密度汎関数理論に基づく電子物性値を、**高速かつ高精度に予測可能**

機械学習を用いた量子化学計算の高速かつ高精度な近似法



- **化合物の何らかの性質を反映する**低次元かつ連続値のベクトル表現を獲得
- 化学・物理学に関する知識・特徴量を一切用いない
  データ駆動型グラフニューラルネットワークの一貫学習を提案
- 密度汎関数理論に基づく電子物性値を、**高速かつ高精度に予測可能**

D1-6: Bayesian Lasso のための実験計画に関する一考察

堀井 俊佑（早大）

※ 発表申込後、やろうと思っていたことと、ほぼ同内容の論文を見つけたので、内容がタイトルと一部合っていません

- スパース線形回帰に対するベイズ的アプローチの分類
  1. 経験ベイズ（Relevance Vector Machine）
  2. スパース性を誘導する事前分布（ラプラス分布など）+ 局所変分近似
  3. **事前分布に階層的な分布を仮定**

$$p(\beta_j|\tau_j) = \mathcal{N}(\beta_j|0, \tau_j^{-1})$$

$$p(\tau_j|a_j) = \mathrm{Exp}(\tau_j|a_j)$$

➡ 周辺分布 $p(\beta_j)$ がラプラス分布

利点

- a_jに事前分布を仮定できる
- 事前知識が無ければ無情報に近い事前分布を

- 逐次実験計画（能動学習）
  - ラベルなしデータが沢山ある状況で、どのデータからラベルを付与するか？
  - ベイズ的アプローチの利点：**情報利得**が最大になるようなデータを選択
    （**ベイズ的逐次実験計画 or ベイズ的能動学習**）

$\boldsymbol{x}_*$ の情報利得：

$$\underbrace{H(p(\boldsymbol{\beta}|\boldsymbol{X}, \boldsymbol{y}))}_{\text{得られているデータの元での事後分布のエントロピー}} - \underbrace{\mathbb{E}_{p(y_*|\boldsymbol{X}, \boldsymbol{y}, \boldsymbol{x}_*)}\left[H(p(\boldsymbol{\beta}|\boldsymbol{X}, \boldsymbol{y}, \boldsymbol{x}_*, y_*)\right]}_{\text{「}\boldsymbol{x}_*\text{に対するラベルを得たもとでの事後分布のエントロピー」の予測分布による期待値}}$$

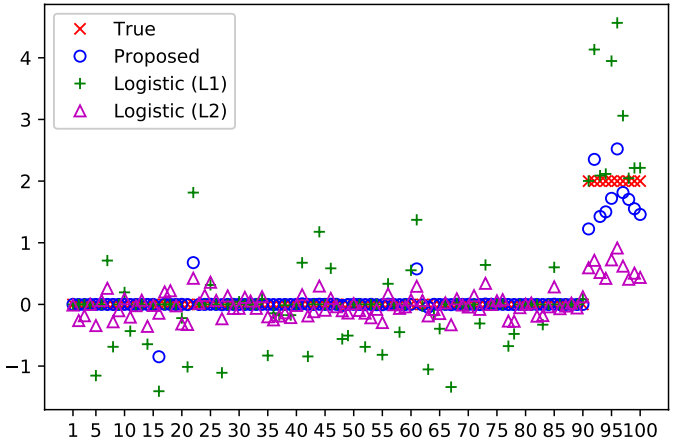
問題　どうやって計算する？

D1-6: Bayesian Lasso のための実験計画に関する一考察

- ベイズ的スパース線形回帰・分類に関する研究

モデル	手法	回帰	分類
ガウシアン事前分布	経験ベイズ	Tipping, 2001	Tipping, 2001
ラプラス分布など	変分ベイズ	Seeger, 2011	Serra, 2016
階層モデル	EM	Figueiredo, 2003	Figueiredo, 2003
	MCMC	Park, 2008	—
	変分ベイズ	Babacan, 2014	本研究

平均場近似+局所変分近似によるアルゴリズム



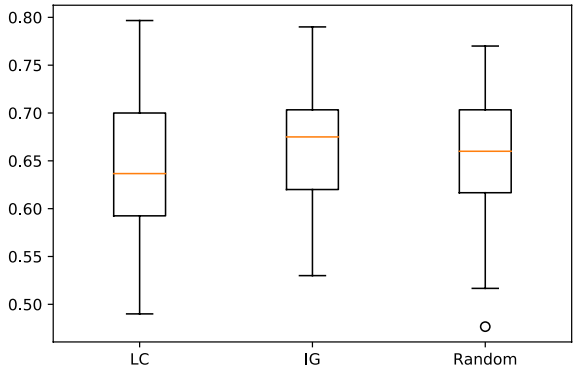
L1正則化ロジスティック回帰よりスパースな解が得られる

- ベイズ的スパース線形回帰・分類と実験計画に関する研究

モデル	回帰	分類
ガウシアン事前分布	Ji, 2008	Silva, 2007*1
ラプラス分布など	Seeger, 2011	—
階層モデル	—*2	本研究

*1: 情報利得は使っていない
*2: 申込み前にやろうと思っていたこと
（[Ji, 2008][Seeger, 2011]とほぼ同じ）

情報利得の近似計算アルゴリズム



従来手法より良い実験計画が得られそう
（ただし計算量：大）

※ 学生プレゼンテーション賞対象

D1-7 高次元特徴量に対するスケーリング法と教師付きスペクトラルクラスタリング

松田 萌望（筑波大学），保國 惠一（筑波大学），櫻井 鉄也（筑波大学，JST/CREST）

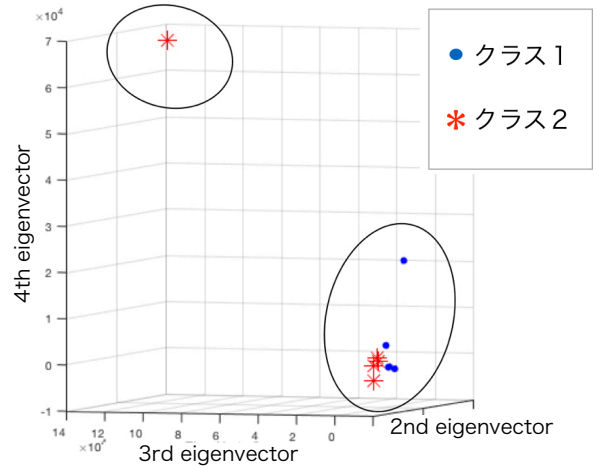
背景

高次元データに対するスペクトラルクラスタリングの適用

所望のクラスタが必ずしも得られるとは限らない

原因１：特徴量のスケールが不揃い
原因２：冗長な特徴量が存在

例：遺伝子発現データ



提案手法

一部のラベル情報を活用し，矩形固有値問題を定式化

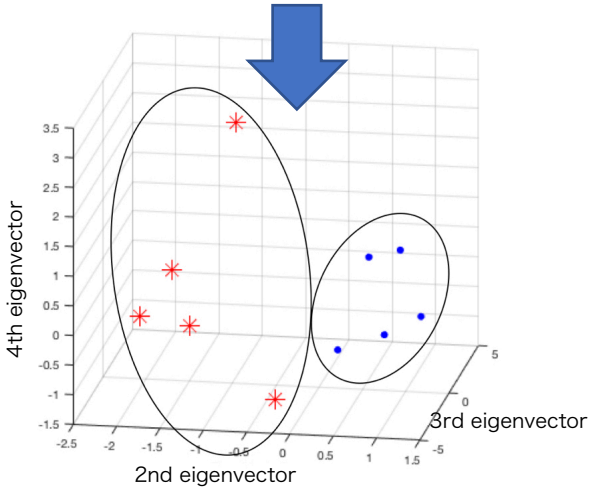
$$\begin{bmatrix} A & \boldsymbol{\alpha} \\ \sum_{i=1}^{n} v_i \hat{\boldsymbol{x}}_i^{\top} & n\sum_{i=1}^{n} v_i \end{bmatrix} \begin{bmatrix} \boldsymbol{s} \\ -1 \end{bmatrix} = \mu \begin{bmatrix} B & \boldsymbol{\beta} \\ \boldsymbol{0}^{\top} & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{s} \\ -1 \end{bmatrix}$$

固有ベクトル $\boldsymbol{s}$：**スケールを調整するための因子**

$$L(S)\boldsymbol{v} = \lambda(S)D(S)\boldsymbol{v}, \quad \boldsymbol{v} \in \mathbb{R}^n \setminus \{\boldsymbol{0}\},$$
$$S = \mathrm{diag}(\boldsymbol{s}), \quad \boldsymbol{v}^{\top} D(S)\boldsymbol{e} = 0, \quad \lambda(S) \in \mathbb{R}$$

データ全体の特徴量のスケールを整える

→ **精度の良い分類が可能に**



※ 学生プレゼンテーション賞対象

D1-7 高次元特徴量に対するスケーリング法と教師付きスペクトラルクラスタリング

松田 萌望（筑波大学），保國 惠一（筑波大学），櫻井 鉄也（筑波大学，JST/CREST）

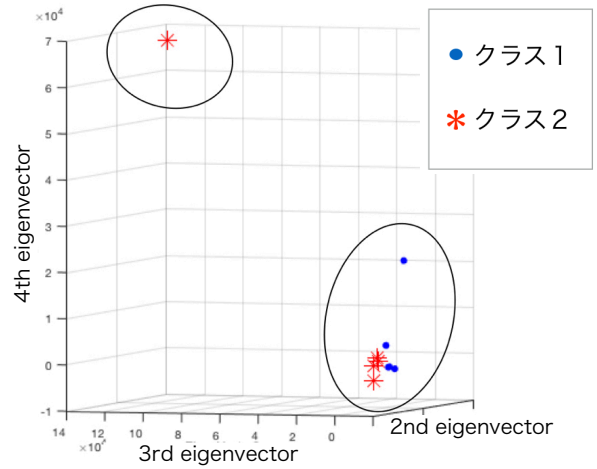
背景

高次元データに対するスペクトラルクラスタリングの適用

所望のクラスタが必ずしも得られるとは限らない

原因１：特徴量のスケールが不揃い
原因２：冗長な特徴量が存在

例：遺伝子発現データ



提案手法

一部のラベル情報を活用し，矩形固有値問題を定式化

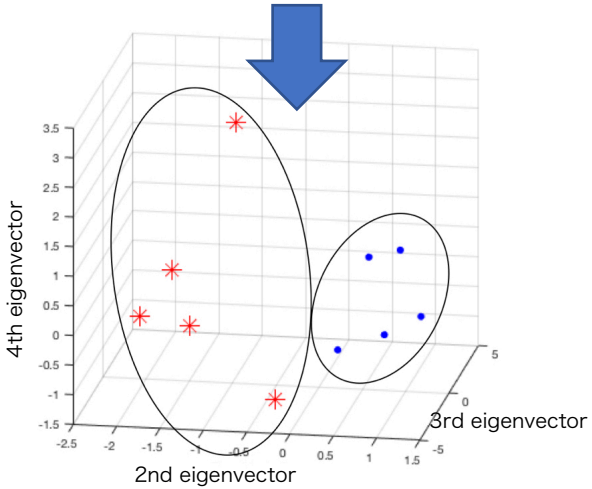
$$\begin{bmatrix} A & \boldsymbol{\alpha} \\ \sum_{i=1}^{n} v_i \hat{\boldsymbol{x}}_i^{\top} & n\sum_{i=1}^{n} v_i \end{bmatrix} \begin{bmatrix} \boldsymbol{s} \\ -1 \end{bmatrix} = \mu \begin{bmatrix} B & \boldsymbol{\beta} \\ \boldsymbol{0}^{\top} & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{s} \\ -1 \end{bmatrix}$$

固有ベクトル $\boldsymbol{s}$：**スケールを調整するための因子**

$$L(S)\boldsymbol{v} = \lambda(S)D(S)\boldsymbol{v}, \quad \boldsymbol{v} \in \mathbb{R}^n \setminus \{\boldsymbol{0}\},$$
$$S = \mathrm{diag}(\boldsymbol{s}), \quad \boldsymbol{v}^{\top} D(S)\boldsymbol{e} = 0, \quad \lambda(S) \in \mathbb{R}$$

データ全体の特徴量のスケールを整える

→ **精度の良い分類が可能に**



D1-8: 因子化情報量基準を用いたLDAのモデル選択

松谷太郎（早稲田大学），浜田道昭（早稲田大学・産総研CBBD-OIL）

- Motivation

  LDAのトピック数選択を高速に行いたい！

- Previous Research

  変分ベイズ法による学習[1]，テンソル分解[2]，etc...

- Method

  周辺化対数尤度をラプラス近似して得られる
  因子化情報量基準（FIC）[3]とShrinkage Mechanismの利用

[1] : Corduneanu, et al. Variational Bayesian Bayesian Model Selection for Mixture Distributions (2001)
[2] : Dehua Cheng, et al. Model Selection for Topic Models via Spectral Decomposition (2015)
[3] : Fujimaki Ryohei, et al. Factorized Asymptotic Bayesian Inference for Mixture Modeling (2012)

D1-8: 因子化情報量基準を用いたLDAのモデル選択

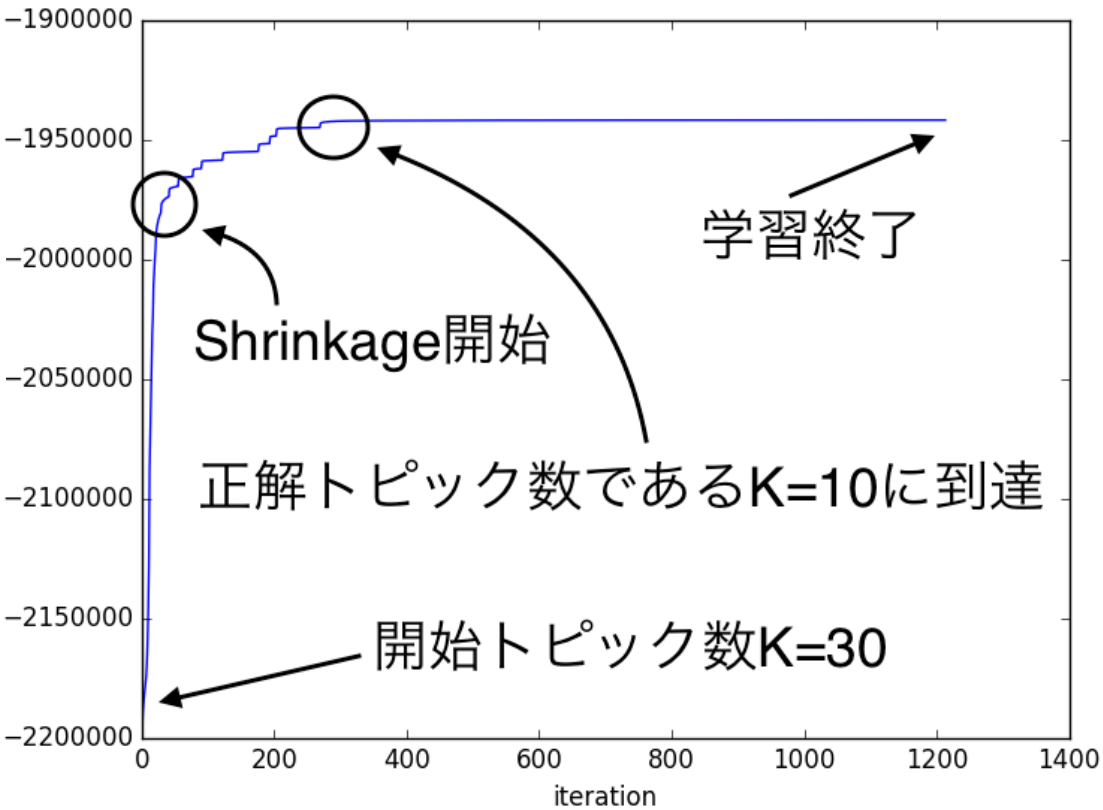
松谷太郎（早稲田大学），浜田道昭（早稲田大学・産総研CBBD-OIL）

・シミュレーション
（がんゲノムの変異データを使用）
文書数 : 500
1文書の単語数 : 1000
語彙数 : 96
正解トピック数 : 10

学習法	トピック数決定に必要な時間
提案手法	7228[s]
VB-LDA	198062[s]

20倍超の高速化！！

学習に伴うFICの推移



D1-9

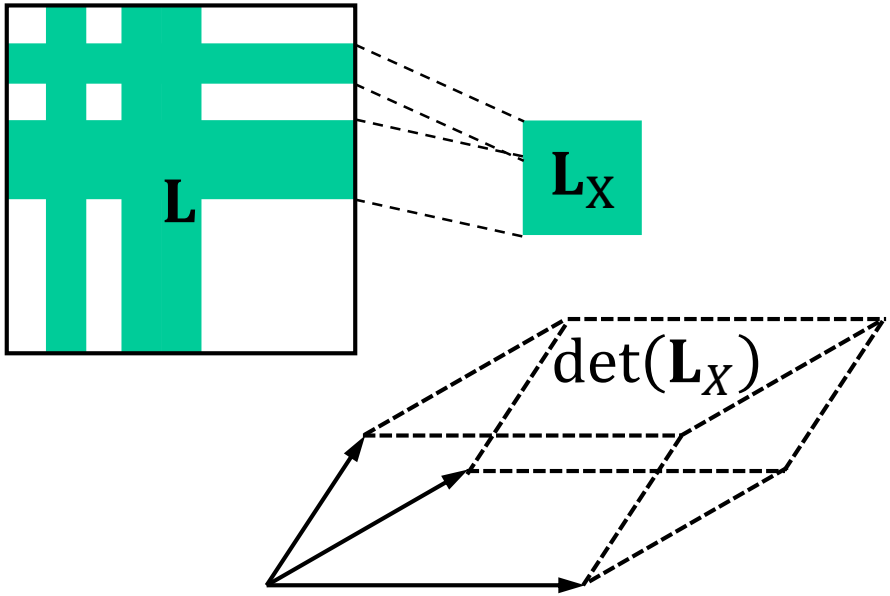
動的行列式点過程

恐神貴行・ルディーレイモンド (IBM東京基礎研究所)， Akshay Goel (Graduate School of Mathematics, Kyushu University)， 白井朋之 (九州大学IMI)， 前原貴憲 (理研AIP)

Determinantal point process (DPP)

$$\mathbb{P}(X) \sim \det(\mathbf{L}_X)$$

Assign high probability to subsets of relevant and diverse items



We propose Dynamic DPP

$$\mathbb{P}(X) \sim \det(\mathbf{L}(t)_X)$$

where kernel changes over time

$$\mathbf{L}(t) = \mathbf{V}(t)\,\mathbf{V}(t)^\top$$

$$\mathbf{V}(t) = \mathbf{B} + \sum_d \mathbf{x}(t-d)\,\mathbf{w}(d)^\top$$

Potential applications:
Sequential recommendation where relevant and diverse items change

This work was supported by JST CREST Grant Number JPMJCR1304, Japan.

D1-9

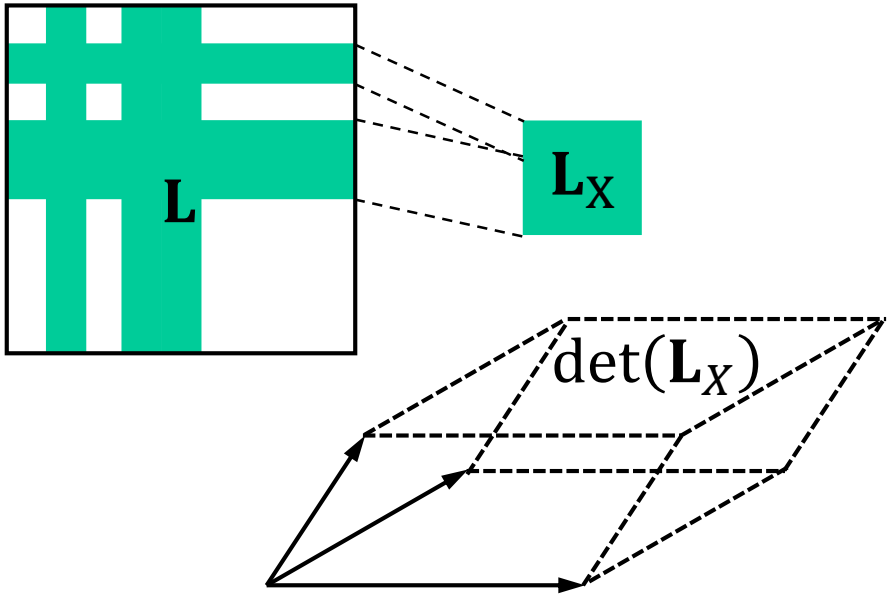
動的行列式点過程

恐神貴行・ルディーレイモンド (IBM東京基礎研究所)， Akshay Goel (Graduate School of Mathematics, Kyushu University)， 白井朋之 (九州大学IMI)， 前原貴憲 (理研AIP)

Determinantal point process (DPP)

$$\mathbb{P}(X) \sim \det(\mathbf{L}_X)$$

Assign high probability to subsets of relevant and diverse items



We propose Dynamic DPP

$$\mathbb{P}(X) \sim \det(\mathbf{L}(t)_X)$$

where kernel changes over time

$$\mathbf{L}(t) = \mathbf{V}(t)\,\mathbf{V}(t)^\top$$

$$\mathbf{V}(t) = \mathbf{B} + \sum_d \mathbf{x}(t-d)\,\mathbf{w}(d)^\top$$

Potential applications:
Sequential recommendation where relevant and diverse items change

This work was supported by JST CREST Grant Number JPMJCR1304, Japan.

小宮山純平（東大） 島尾尭（Purdue University）

D1-10: 2ステージ回帰による公正配慮型学習

- (s, x, y): xからyの予測$\hat{y}$を出力（＝分類、回帰）
- 例：s=性別/年齢, x=特徴量, y=年収
- 公平性： $\Pr[\hat{y}|s] = \Pr[\hat{y}]$
  - 性別・年齢によって年収に差があってはいけない
  - xからyを学習すると…xとsの相関を暗に学習してしまい、差別が生じる
- 2ステージ回帰：xのうちsから予測できる部分を差し引いた$\hat{x}$を作り（ステージ1）、$\hat{x}$からyを学習（ステージ2）
  - 公平性を実現

小宮山純平（東大） 島尾尭（Purdue University）

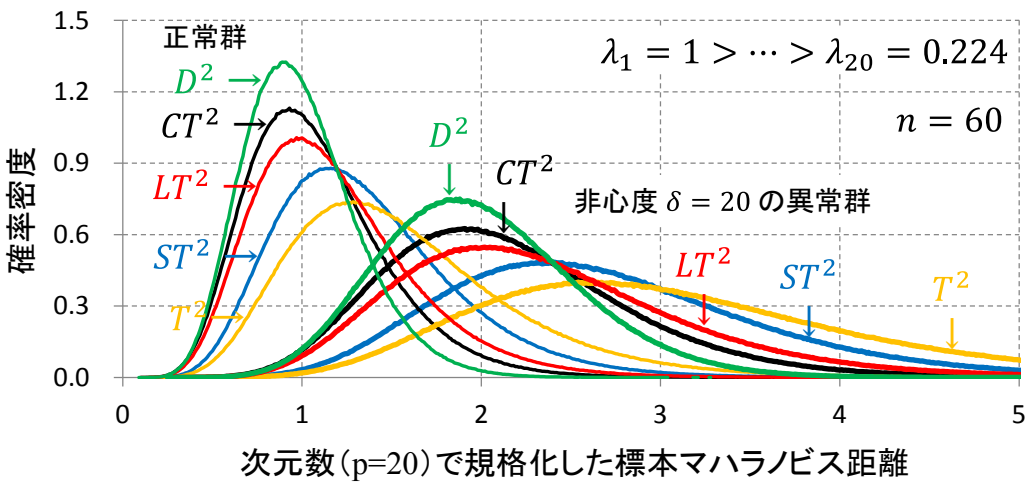
D1-10: 2ステージ回帰による公正配慮型学習

- (s, x, y): xからyの予測$\hat{y}$を出力（＝分類、回帰）
- 例：s=性別/年齢, x=特徴量, y=年収
- 公平性： $\Pr[\hat{y}|s] = \Pr[\hat{y}]$
  - 性別・年齢によって年収に差があってはいけない
  - xからyを学習すると…xとsの相関を暗に学習してしまい、差別が生じる
- 2ステージ回帰：xのうちsから予測できる部分を差し引いた$\hat{x}$を作り（ステージ1）、$\hat{x}$からyを学習（ステージ2）
  - 公平性を実現

D1-11　母固有値の推定を必要としない標本マハラノビス距離補正法の改良

小林 靖之（帝京大学 理工学部）

- 標本マハラノビス距離 $T^2 = (y-\bar{x})'S^{-1}(y-\bar{x})$ は学習サンプル数nの不足により識別性能が劣化するため、有限nのT^2から$n \to \infty$に当たる母マハラノビス距離 D^2の推定が望ましい。
- そこで、真の共分散行列 Σ の固有値（母固有値 λ）の正確な推定を前提として、T^2をD^2へ補正する方法（図のCT^2）を考案し、良好な結果を得た。
- しかし、CT^2に必要な母固有値λの正確な推定は実用上難しい。そこで、 母固有値λの推定なしでT^2をD^2へ補正する方法（図のST^2）を考案したが、CT^2よりもST^2は劣っていた。
- 今回、ほぼ同様な補正式でありながら、母固有値λの推定なしでST^2よりもD^2への補正性能が優れる改良方法（図のLT^2）を考案した。その概要と検討経過を報告するので、議論したい。

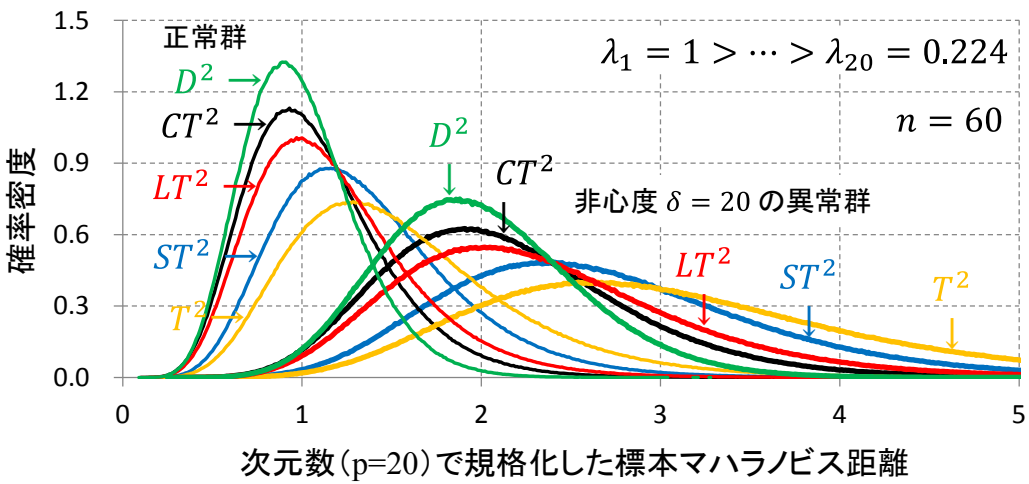


謝辞：
この研究の一部は
JSPS科研費15H02798
の助成を受けた。

D1-11　母固有値の推定を必要としない標本マハラノビス距離補正法の改良

小林 靖之（帝京大学 理工学部）

- 標本マハラノビス距離 $T^2 = (y-\bar{x})'S^{-1}(y-\bar{x})$ は学習サンプル数nの不足により識別性能が劣化するため、有限nのT^2から$n \to \infty$に当たる母マハラノビス距離 D^2の推定が望ましい。
- そこで、真の共分散行列 Σ の固有値（母固有値 λ）の正確な推定を前提として、T^2をD^2へ補正する方法（図のCT^2）を考案し、良好な結果を得た。
- しかし、CT^2に必要な母固有値λの正確な推定は実用上難しい。そこで、 母固有値λの推定なしでT^2をD^2へ補正する方法（図のST^2）を考案したが、CT^2よりもST^2は劣っていた。
- 今回、ほぼ同様な補正式でありながら、母固有値λの推定なしでST^2よりもD^2への補正性能が優れる改良方法（図のLT^2）を考案した。その概要と検討経過を報告するので、議論したい。



謝辞：
この研究の一部は
JSPS科研費15H02798
の助成を受けた。

D1-12

Robust and Sparse Online GLM

川島　孝行 (総研大)　藤澤　洋徳 (統数研)

gamma-cross entropy

$$d_\gamma(g(y|x), f(y|x); g(x)) = -\frac{1}{\gamma}\log E_{g(x,y)}\left[\frac{f(y|x)^\gamma}{\left(\int f(y|x)^{1+\gamma}dy\right)^{\frac{\gamma}{1+\gamma}}}\right].$$

Regularized expected risk minimization

$$\underset{\theta}{\operatorname{argmin}} -\frac{1}{\gamma}\log E_{g(x,y)}\left[\frac{f(y|x;\theta)^\gamma}{\left(\int f(y|x;\theta)^{1+\gamma}dy\right)^{\frac{\gamma}{1+\gamma}}}\right] + \lambda P(\theta).$$

Regularized empirical risk minimization

$$\underset{\theta}{\operatorname{argmin}} -\frac{1}{\gamma}\log \frac{1}{n}\sum_{i=1}^{n}\frac{f(y_i|x_i;\theta)^\gamma}{\left(\int f(y|x_i;\theta)^{1+\gamma}dy\right)^{\frac{\gamma}{1+\gamma}}} + \lambda P(\theta).$$

ただし，Pは正則化項.

例）f(y|x;θ)が**正規分布**　→　線形回帰
二項分布　→　ロジスティック回帰
ポアソン分布　→　ポアソン回帰

学生優秀プレゼンテーション賞対象

Robust and Sparse Online GLM

川島　孝行 (総研大)　藤澤　洋徳 (統数研)

Regularized empirical risk minimization

$$\underset{\theta}{\operatorname{argmin}} -\frac{1}{\gamma}\log\frac{1}{n}\sum_{i=1}^{n}\frac{f(y_i|x_i;\theta)^{\gamma}}{\left(\int f(y|x_i;\theta)^{1+\gamma}dy\right)^{\frac{\gamma}{1+\gamma}}}+\lambda P(\theta).$$

問題：　ポアソン回帰の場合に，$\left(\int f(y|x_i;\theta)^{1+\gamma}dy\right)^{\frac{\gamma}{1+\gamma}}$ が**一般化超幾何級数**.

→パラメータに関して**陽**に解けない.

解決法：　本来の目的である，**Regularized expected risk minimization**の方を解く.

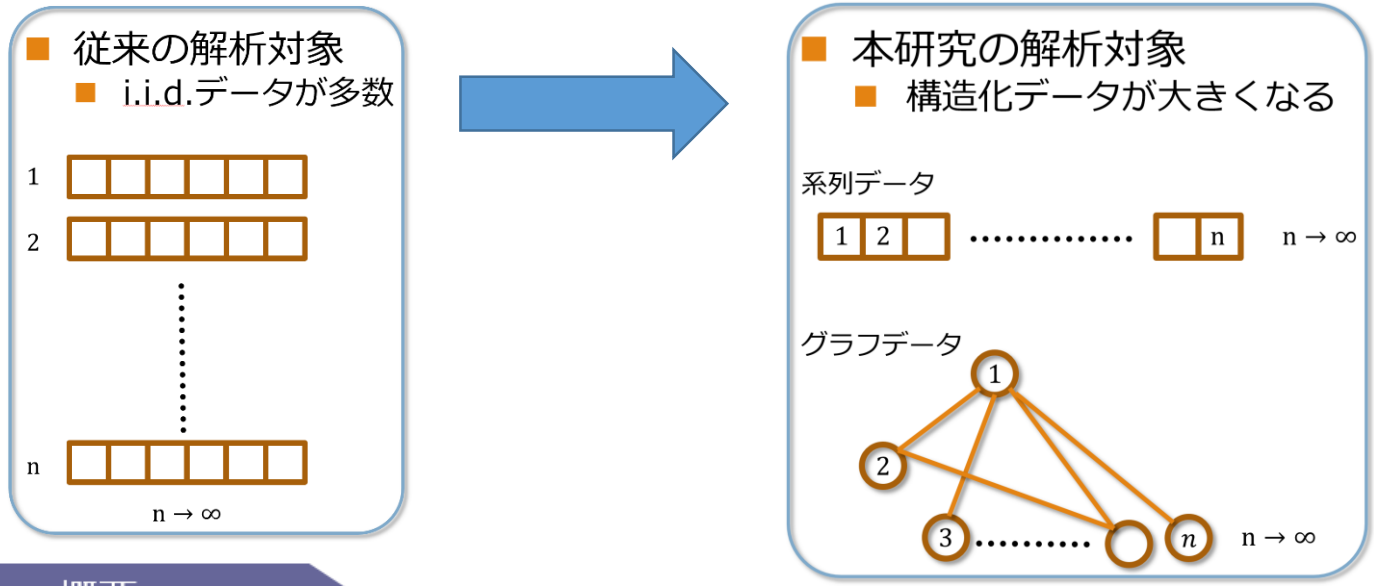
→**確率的最適化**の導入.

結果：　**大域的収束性**を保証した**オンライン**推定アルゴリズム.
大規模なデータにも対応可.
ポアソン回帰だけでなく，線形回帰，ロジスティック回帰も
同時に**オンライン**化.

学生優秀プレゼンテーション賞対象

D1-13 構造化データにおける ベイズクラスタリング精度について

山崎啓介, 川本達郎（産総研AI研究センター）



概要

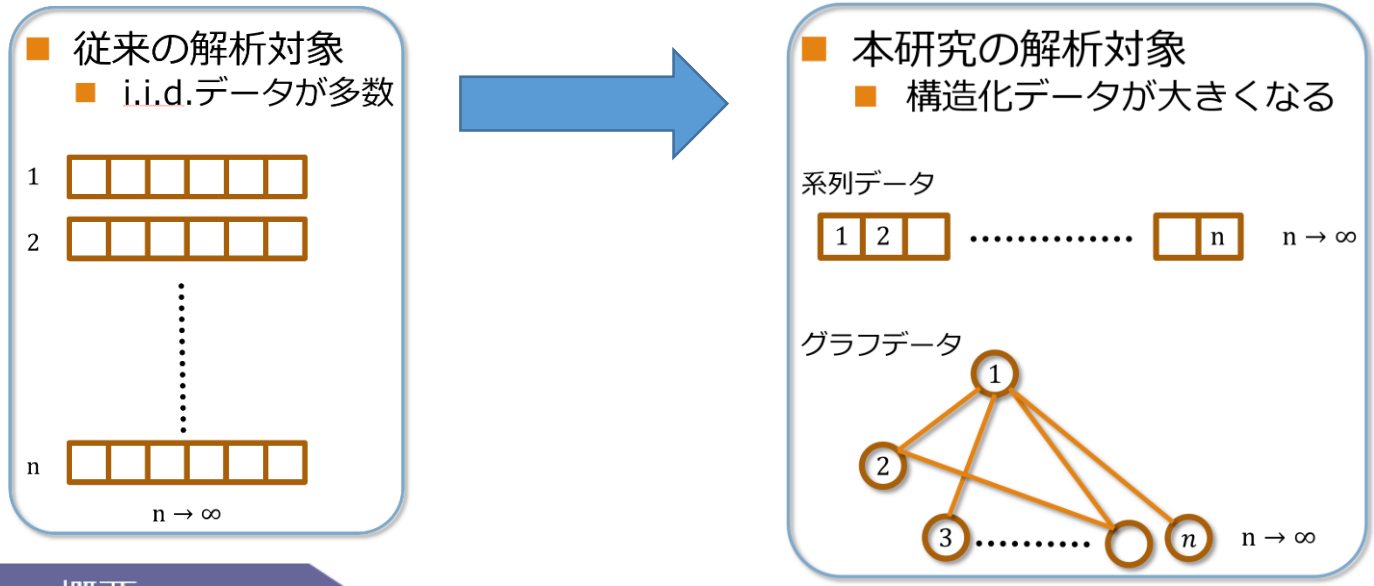
背景： クラスタリング問題をラベル分布の推定としてKL divergenceで精度を定式化
i.i.d.データの数が十分大きい場合に漸近評価が可能

問題： 系列データやグラフデータはi.i.d.ではないため精度が不明

結果： 系列長が十分長い場合、グラフノード数が十分大きい場合の精度の上界を導出
隠れマルコフモデルと確率的ブロックモデル＋共役事前分布
モデルサイズに冗長性があるとき、MCMC法には刈込作用がある

D1-13 構造化データにおける ベイズクラスタリング精度について

山崎啓介, 川本達郎（産総研AI研究センター）



概要

背景： クラスタリング問題をラベル分布の推定としてKL divergenceで精度を定式化
i.i.d.データの数が十分大きい場合に漸近評価が可能

問題： 系列データやグラフデータはi.i.d.ではないため精度が不明

結果： 系列長が十分長い場合、グラフノード数が十分大きい場合の精度の上界を導出
隠れマルコフモデルと確率的ブロックモデル＋共役事前分布
モデルサイズに冗長性があるとき、MCMC法には刈込作用がある

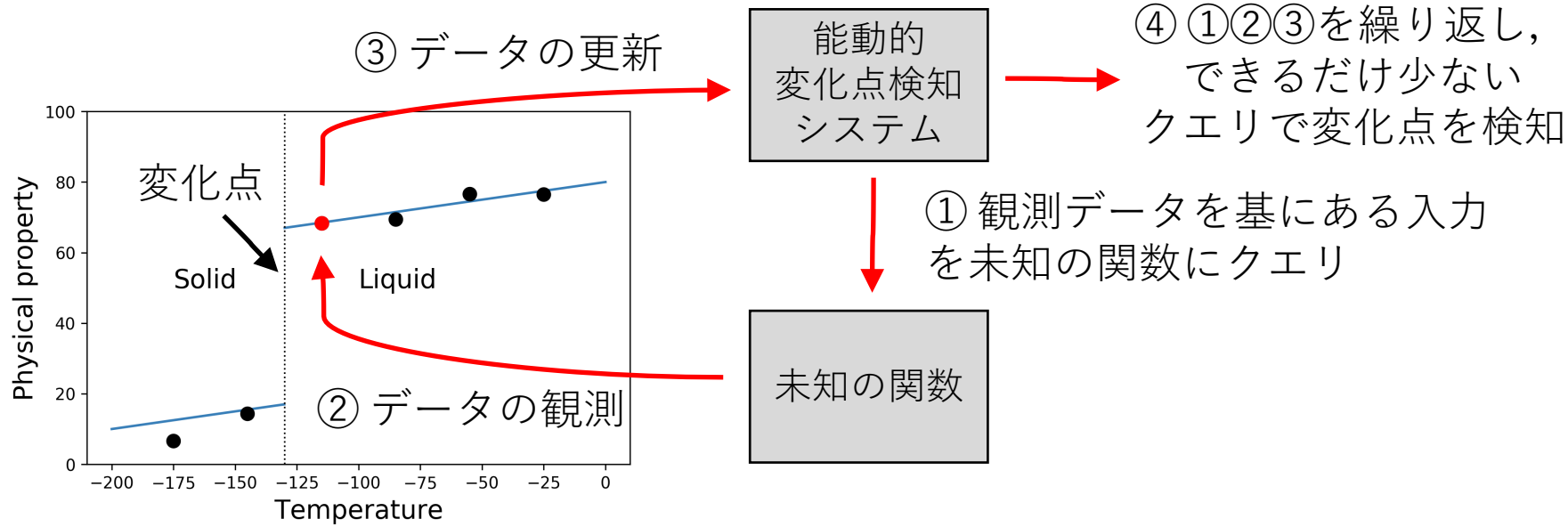
能動的変化点検知

発表ID：D1-14

林勝悟*　河原吉伸**　鹿島久嗣*
*京都大学　**大阪大学

既存の変化点検知では考慮されていなかったデータ取得にコストを要する能動的変化点検知を定式化し，汎用的な枠組みを提案する

能動的変化点検知の流れ



応用例：材料科学における，ある物質の物性（抵抗，比熱等）が大きく変動する相転移温度の実験的同定

未知の関数：実験系，入力：温度，出力：物性，コスト：時間

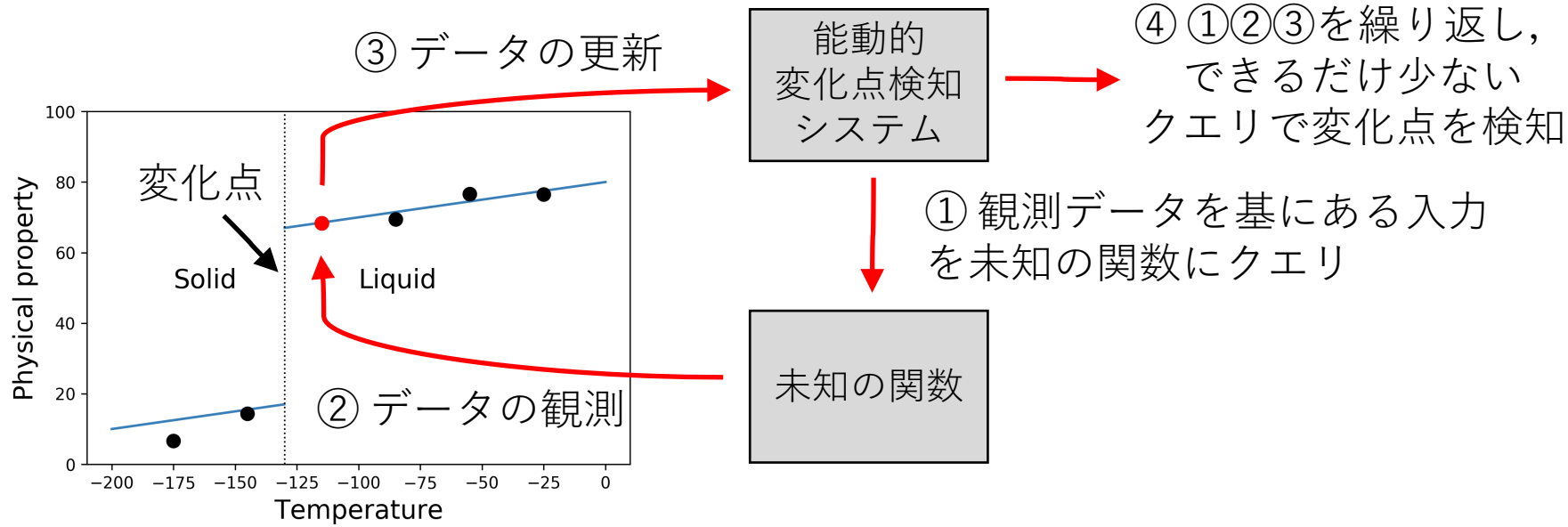
能動的変化点検知

発表ID：D1-14

林勝悟*　河原吉伸**　鹿島久嗣*
*京都大学　**大阪大学

既存の変化点検知では考慮されていなかったデータ取得にコストを要する能動的変化点検知を定式化し，汎用的な枠組みを提案する

能動的変化点検知の流れ



応用例：材料科学における，ある物質の物性（抵抗，比熱等）が大きく変動する相転移温度の実験的同定

未知の関数：実験系，入力：温度，出力：物性，コスト：時間

D1-15

ソフトラベルの非確率的解釈と分類学習への応用

○高瀬　朝海[1], 小山　聡[1,2], 栗原　正仁[1]

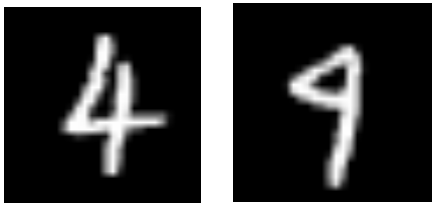
1: 北海道大学, 2: 理研AIP

データとクラスの関係を[0, 1]の実数値で表すソフトラベル（⇔{0, 1}のハードラベル）は、分類の難しいデータを上手く表現することができる。

ソフトラベルの2通りの解釈

1. 確率として　・・・　統計的機械学習におけるクラスの事後確率
2. 帰属度として　・・・　クラスにどの程度属しているかの度合い

クラウドソーシングでソフトラベルを付ける場合、ワーカにとって、帰属度、すなわち各クラスの模範的なサンプルとの類似度を答える方が、直感的なタスクといえる。



MNISTデータセット
（ラベル4のサンプル）

・4である確率・・・？

・どれくらい4らしいか？

フィッシャーの線形判別をソフトラベル用に拡張し、クラウドソーシングを利用して作られた、各解釈に基づいたソフトラベルを用いて実験を行う。

D1-15

ソフトラベルの非確率的解釈と分類学習への応用

○高瀬　朝海[1], 小山　聡[1,2], 栗原　正仁[1]

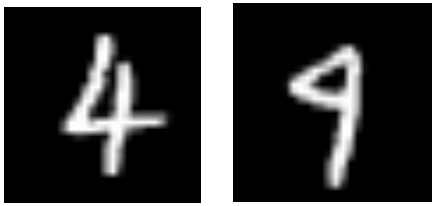
1: 北海道大学, 2: 理研AIP

データとクラスの関係を[0, 1]の実数値で表すソフトラベル（⇔{0, 1}のハードラベル）は、分類の難しいデータを上手く表現することができる。

ソフトラベルの2通りの解釈

1. 確率として　・・・　統計的機械学習におけるクラスの事後確率
2. 帰属度として　・・・　クラスにどの程度属しているかの度合い

クラウドソーシングでソフトラベルを付ける場合、ワーカにとって、帰属度、すなわち各クラスの模範的なサンプルとの類似度を答える方が、直感的なタスクといえる。



MNISTデータセット
（ラベル4のサンプル）

・4である確率・・・？

・どれくらい4らしいか？

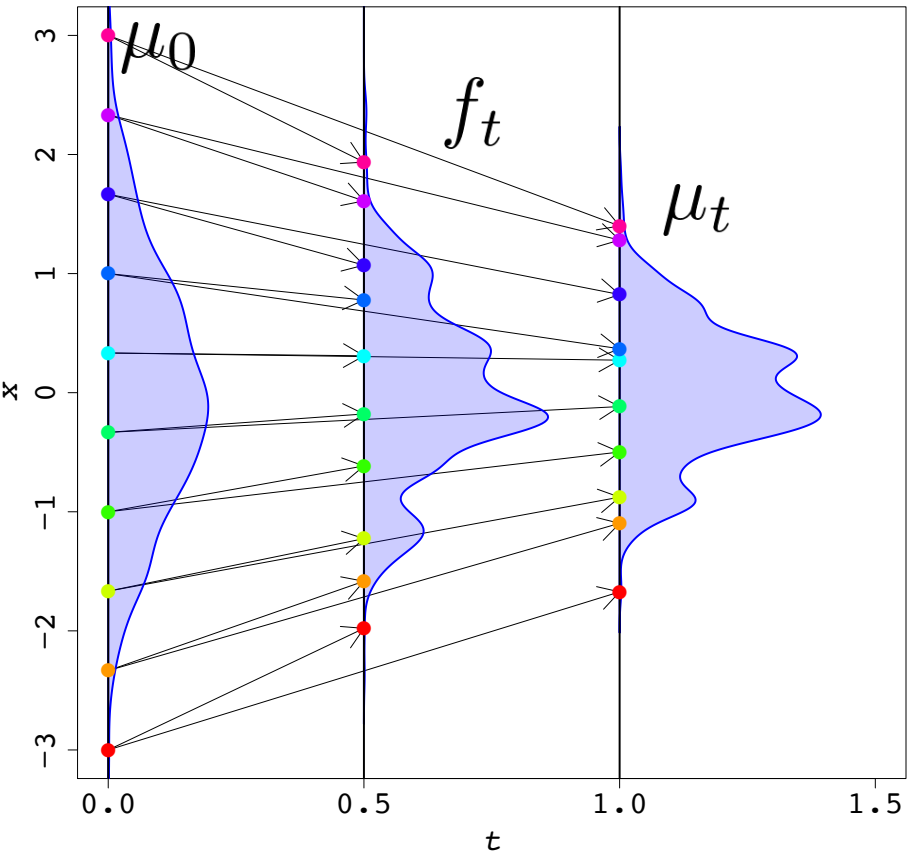
フィッシャーの線形判別をソフトラベル用に拡張し、クラウドソーシングを利用して作られた、各解釈に基づいたソフトラベルを用いて実験を行う。

D1-16

一般の雑音分布による Denoising Autoencoder の輸送解析

早稲田大学　園田翔　村田昇

DAE学習則 $$f(x+\varepsilon)=x, \quad x\sim\mu_0,\ \varepsilon\sim\nu(0,tI)$$



学習済DAEは輸送写像とみなせる

$$f_t(x)=x-\frac{\int\varepsilon\mu_0(\varepsilon-x)\nu_t(\varepsilon)\mathrm{d}\varepsilon}{\int\mu_0(\varepsilon-x)\nu_t(\varepsilon)\mathrm{d}\varepsilon}$$

雑音分布が誘導する輸送の性質を調べたい

$$\frac{\mathrm{d}}{\mathrm{d}t}\mu_t=-\mathrm{grad}\mathcal{F}_\nu[\mu_t]$$

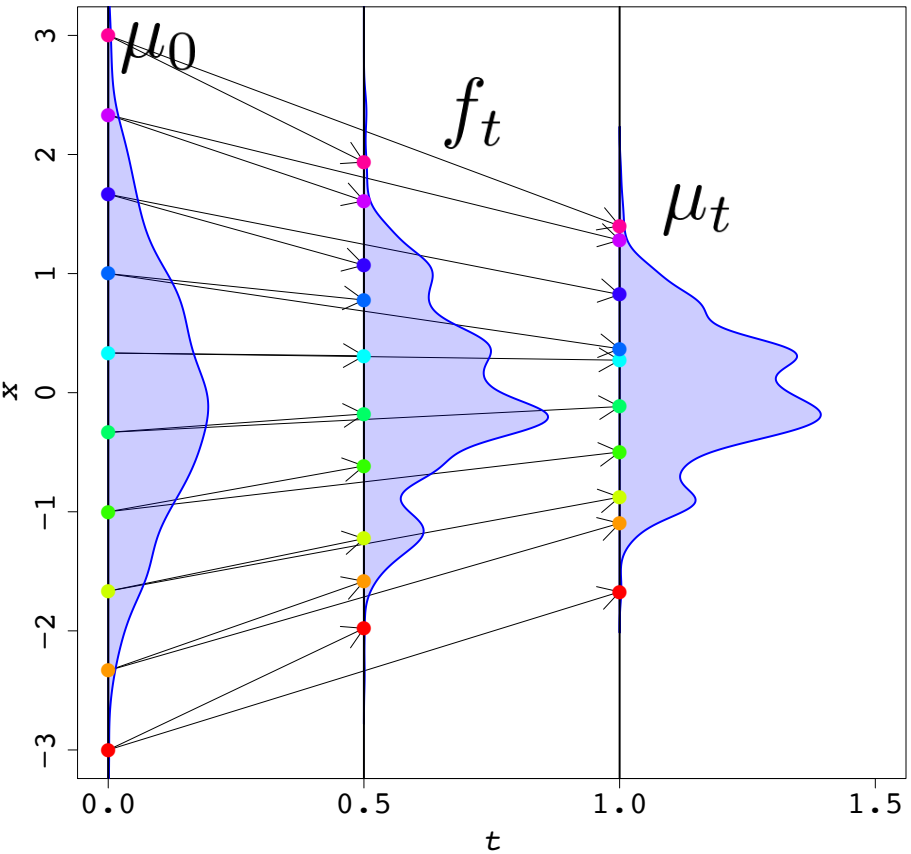
キーワード:
平均の推定量，Wasserstein勾配流，連続方程式
q-Gaussian，Tsallisエントロピー

D1-16

一般の雑音分布による Denoising Autoencoder の輸送解析

早稲田大学　園田翔　村田昇

DAE学習則 $$f(x+\varepsilon)=x, \quad x\sim\mu_0,\ \varepsilon\sim\nu(0,tI)$$



学習済DAEは輸送写像とみなせる

$$f_t(x)=x-\frac{\int\varepsilon\mu_0(\varepsilon-x)\nu_t(\varepsilon)\mathrm{d}\varepsilon}{\int\mu_0(\varepsilon-x)\nu_t(\varepsilon)\mathrm{d}\varepsilon}$$

雑音分布が誘導する輸送の性質を調べたい

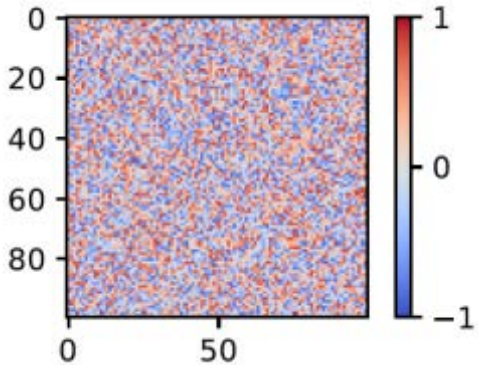
$$\frac{\mathrm{d}}{\mathrm{d}t}\mu_t=-\mathrm{grad}\mathcal{F}_\nu[\mu_t]$$

キーワード:
平均の推定量，Wasserstein勾配流，連続方程式
q-Gaussian，Tsallisエントロピー

[D1-17] 深層学習のための線形制約付きウェイト

沓名 拓郎（豊田中央研究所）

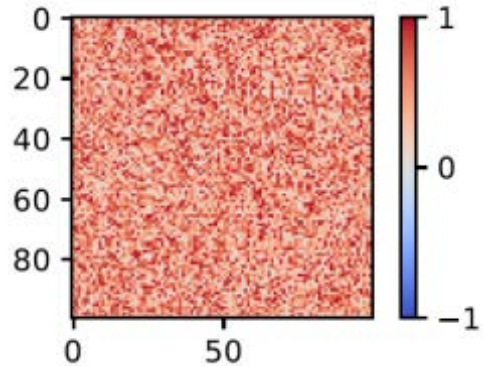
$W \in R^{100\times100}$



ランダム行列

×

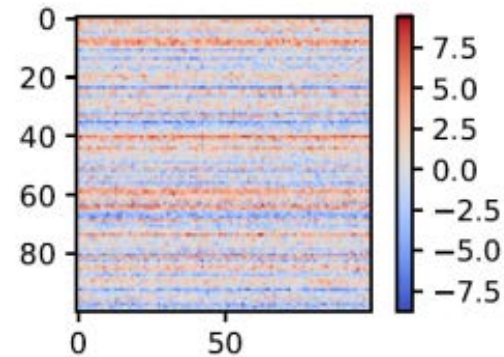
$A \in R^{100\times100}$



ランダム行列

＝

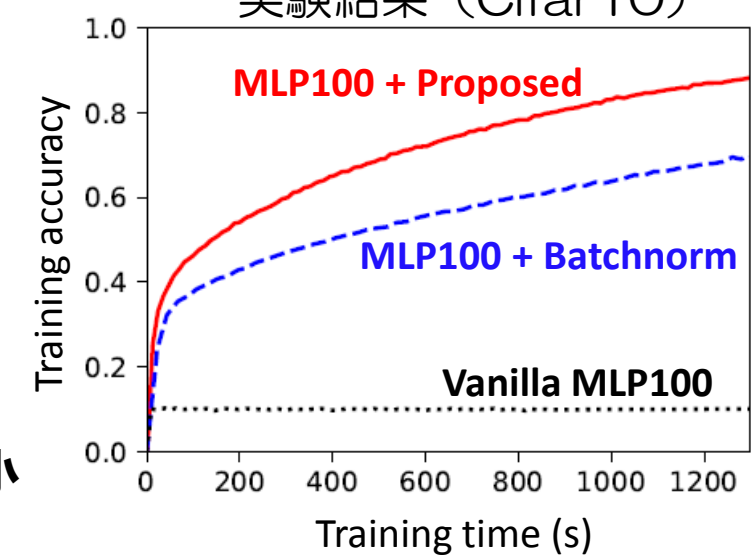
WA



縞模様！？

- ディープネット内部でも同様の偏り（縞模様）が発生
  → 勾配消失などの要因
- <u>偏り（縞模様）を防ぐ方法</u>を提案
  各ウェイトベクトルに対し「総和がゼロ」と制約
  → 100層のMLPが学習可能に
- バッチ正規化もこの「偏り」を防いでいると見なせる
  → 提案手法の方が収束速い＆計算オーバーヘッド小

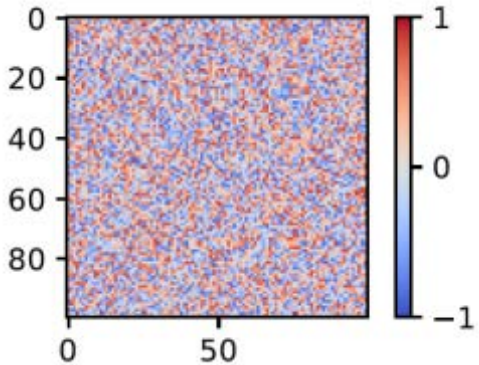
実験結果（Cifar10）



[D1-17] 深層学習のための線形制約付きウェイト

沓名 拓郎（豊田中央研究所）

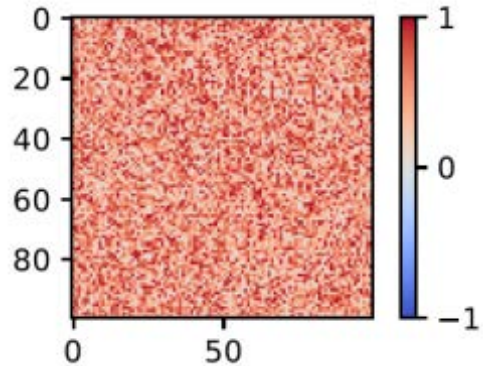
$W \in R^{100\times100}$



ランダム行列

×

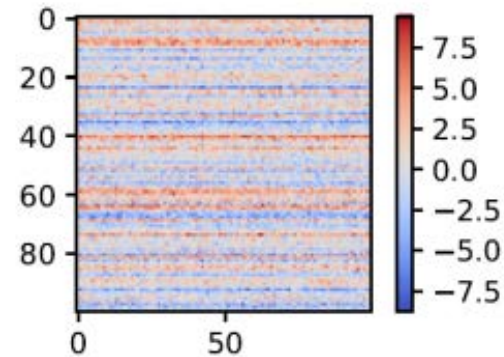
$A \in R^{100\times100}$



ランダム行列

＝

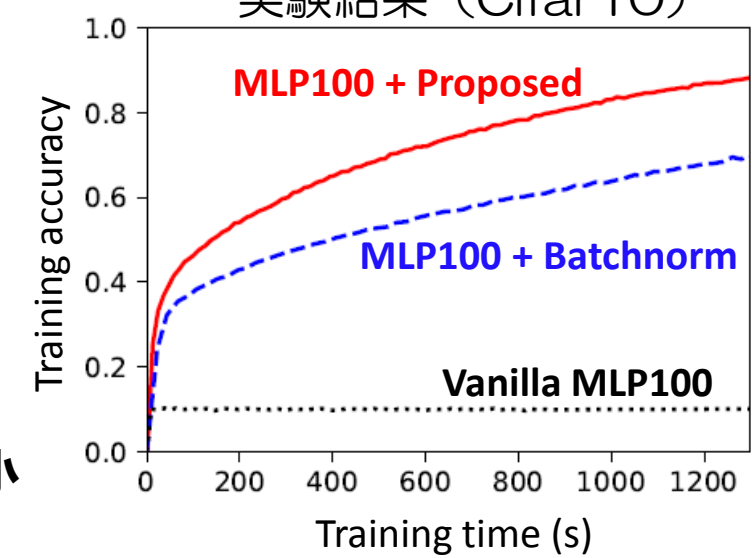
WA



縞模様！？

- ディープネット内部でも同様の偏り（縞模様）が発生
  → 勾配消失などの要因
- <u>偏り（縞模様）を防ぐ方法</u>を提案
  各ウェイトベクトルに対し「総和がゼロ」と制約
  → 100層のMLPが学習可能に
- バッチ正規化もこの「偏り」を防いでいると見なせる
  → 提案手法の方が収束速い＆計算オーバーヘッド小

実験結果（Cifar10）



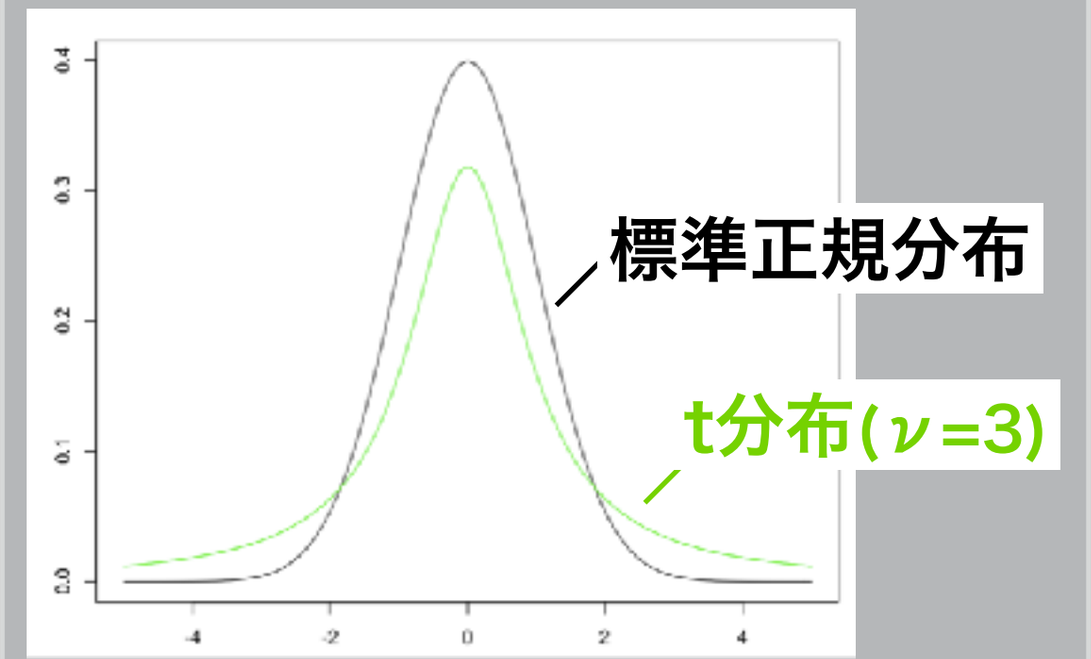
D1-18

学生優秀プレゼンテーション賞対象

テンソルt分布とその応用

大阪大学　宇野光平

t分布は正規分布よりロバスト



各分布の確率密度関数

t分布を用いたデータ解析手法

混合モデル(Peel & McLachlan)
グラフィカルモデル(Finegold & Drton)
混合エキスパート(Chamroukhi)
因子分析(Zhang et al.)
など

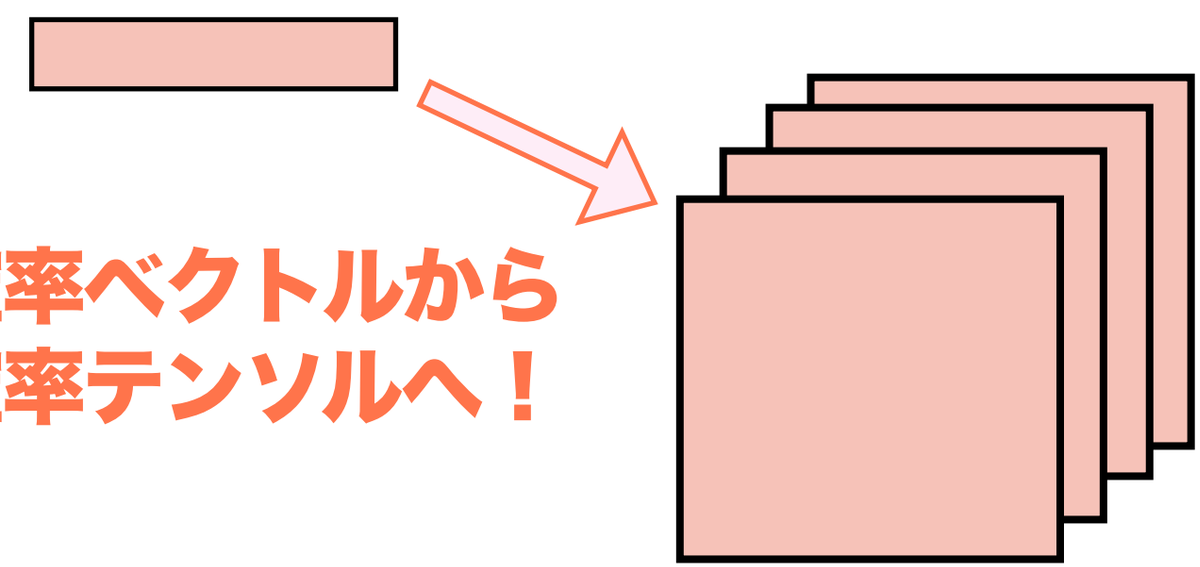
t分布を用いたデータ解析手法はiidを仮定…

成り立たない場合もあるのでは？🤔

- 同じ個体の別臓器での遺伝子発現
- 同じクラスでの時期別テスト得点

など

テンソル(多相)構造に対応するために



確率ベクトルから確率テンソルへ！

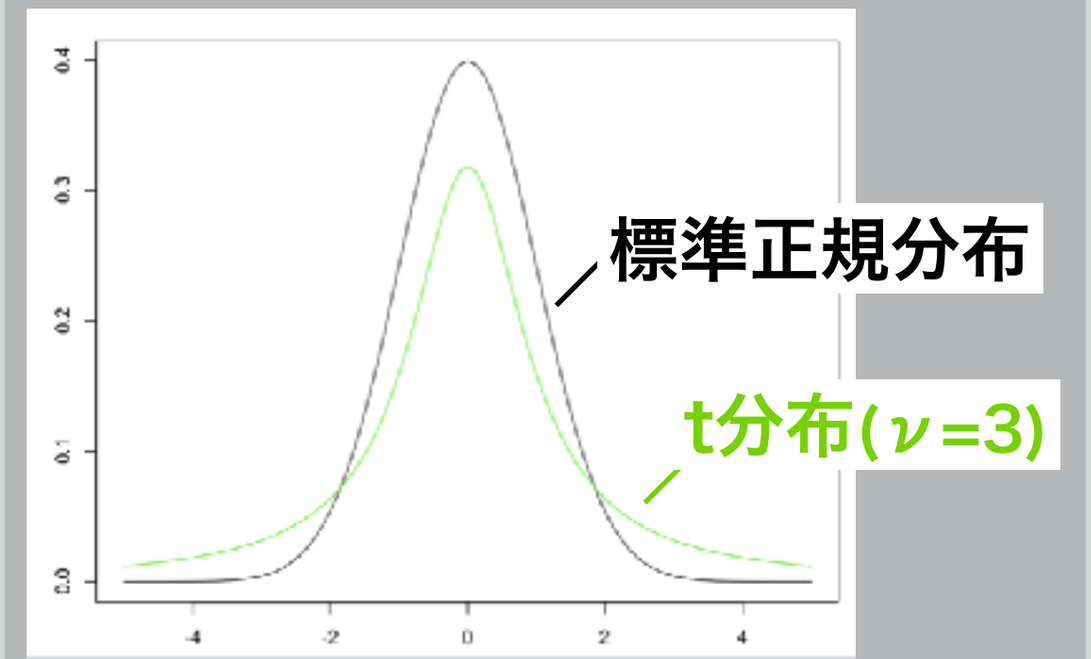
D1-18

学生優秀プレゼンテーション賞対象

テンソルt分布とその応用

大阪大学　宇野光平

t分布は正規分布よりロバスト



各分布の確率密度関数

t分布を用いたデータ解析手法

混合モデル(Peel & McLachlan)
グラフィカルモデル(Finegold & Drton)
混合エキスパート(Chamroukhi)
因子分析(Zhang et al.)
など

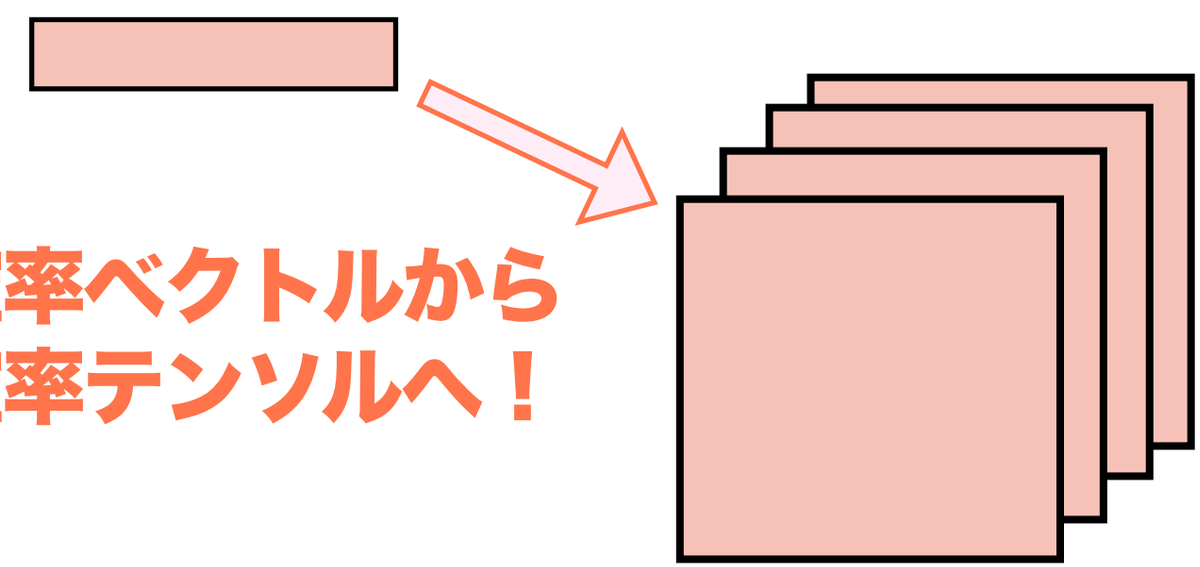
t分布を用いたデータ解析手法はiidを仮定…

成り立たない場合もあるのでは？🤔

- 同じ個体の別臓器での遺伝子発現
- 同じクラスでの時期別テスト得点

など

テンソル(多相)構造に対応するために



確率ベクトルから確率テンソルへ！

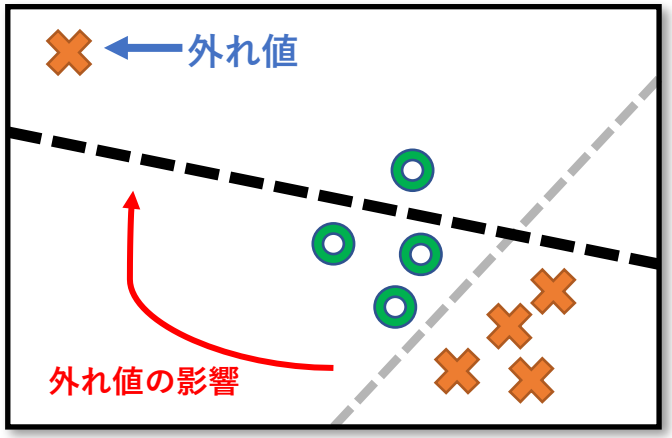
D1-19

学生優秀プレゼンテーション賞対象

ベータダイバージェンスと逐次最小二乗法を用いたロバストな二値分類

土田 雄輝*, 福永 修一*
*東京都立産業技術高等専門学校

- 二値分類のオンライン学習
  - 既存手法は**外れ値に弱い**



外れ値が含まれる二値分類

- 提案手法

推定のロバスト化に有効

ベータダイバージェンス
(Basu et al. 1998,
Eguchi and Kano 2001)
最小化

＋

逐次
最小二乗法

→

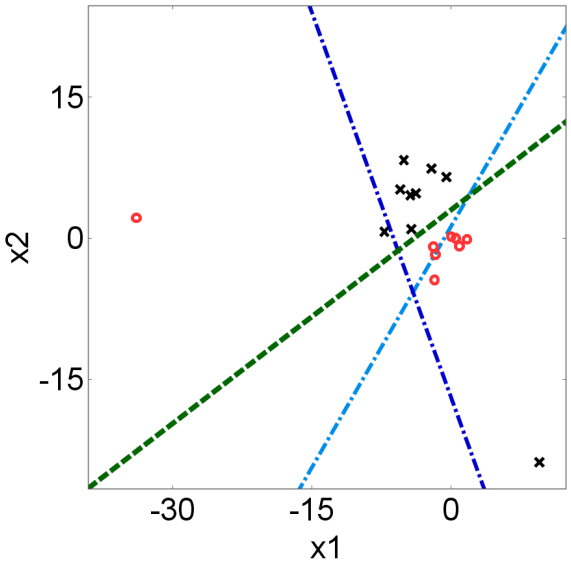
外れ値に強い
逐次型二値分類器

D1-19 ベータダイバージェンスと逐次最小二乗法を用いたロバストな二値分類

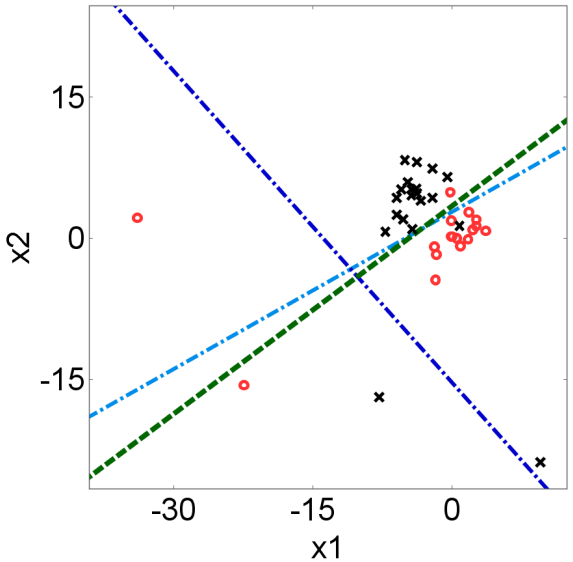
土田 雄輝, 福永 修一(東京都立産業技術高等専門学校)

■ 外れ値を含んだデータに対する推定結果

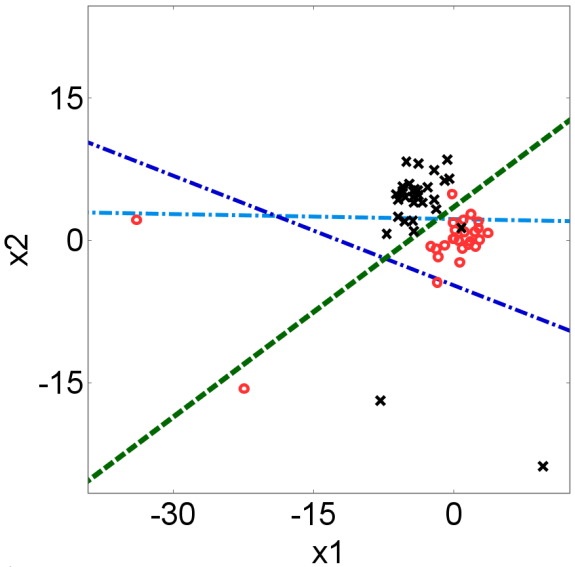
提案手法と従来手法(**AROW**, **SCW**)の二値分類結果を比較



学習データ数20個



学習データ数40個



学習データ数60個

- **AROW(青線)**, **SCW(水色線)**：外れ値の影響を受けている
- **提案手法(緑線)**：外れ値の影響を受けていない

➡ **提案手法は従来手法より推定精度が高い**

D1-20

学生優秀プレゼンテーション賞対象

ベータダイバージェンスを用いたロバスト部分最小二乗法

藤野 眞人*, 福永 修一*
*東京都立産業技術高等専門学校

■ 部分最小二乗法

- 説明変数を低次元空間に射影し,
  得られた潜在変数から目的変数を推定
- 外れ値が混入すると推定に影響

ベータダイバージェンス(Eguchi & Kano 2001, Basu et al. 1998)
最小化によるロバスト化が有効

↓

ベータダイバージェンスを用いた
ロバスト部分最小二乗法を提案

ベータダイバージェンスを用いたロバスト部分最小二乗法

藤野 眞人*, 福永 修一*
*東京都立産業技術高等専門学校

■ 従来手法との比較

- 数値例の比較から提案手法の有効性を検証
- 従来手法
  - ➢ Partial Robust M-Regression(Serneels et al. 2005)

表1. 目的変数モデルと推定値との平均二乗誤差(MSE)

手法	部分最小二乗法	Partial Robust M-Regression	提案手法
MSE	10.5	0.214	**0.129**

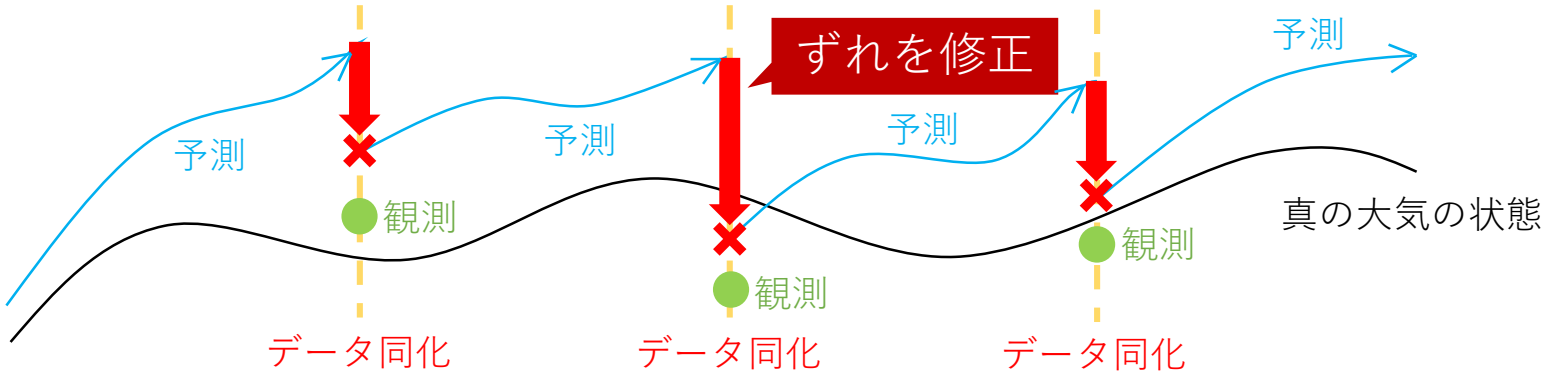
提案手法は従来手法より平均二乗誤差が小さい

D1-21

気象シミュレーションにおける衛星画像データ同化への機械学習の応用

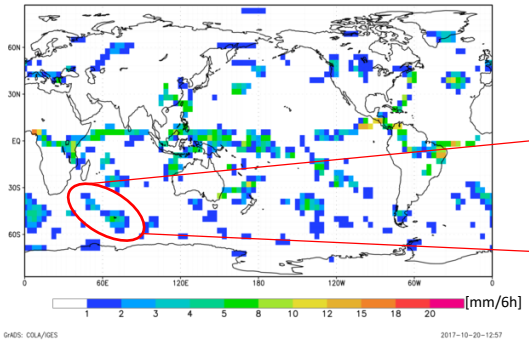
粟津妙華，大塚成徳，三好建正（理化学研究所　計算科学研究機構）

- **データ同化**
  - シミュレーション(数理モデル)と現実世界とのずれを、観測によって修正する。
  - 主に気象分野で発展
    - 気象シミュレーションに気象衛星や気球観測による観測データを同化することにより、予報を改善。
      - ➢ 気象衛星：赤外などの画像データ
      - ➢ 気球観測：気温、気圧、湿度、風

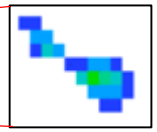


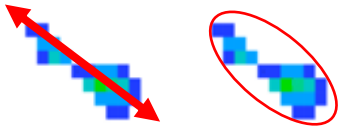
降水領域の形状特徴量を用いたデータ同化

低解像度全球大気モデル



降水領域



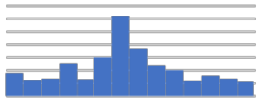


傾き　　形状

一般的な幾何形状特徴量

確率密度関数

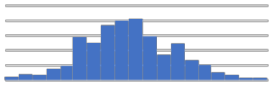
傾き特徴量



トレーニングデータ：80,000枚
テストデータ：10,000枚

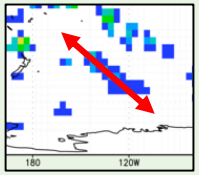
オートエンコーダによる特徴量の抽出

オートエンコーダによって抽出した特徴量

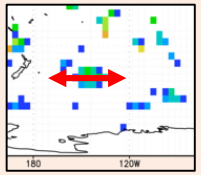


例　傾き特徴量の同化

同化前

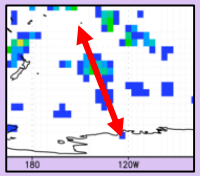


観測

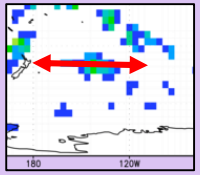


同化後

コントロール実験



傾き＋コントロール実験



コントロール実験: 東西風，南北風，気温，水蒸気量を同化

目的

降水領域の形状特徴量を用いたデータ同化実験を行い、降水領域の形状を改善することで、気象場全体の改善をもたらし、トータルな予測を改善する。

結果と考察

一般的な幾何形状特徴量の同化実験を1時刻分行った。結果、降水領域の形状は変化し、観測に近づいた。このことから、形状特徴量の同化は、降水領域の改善に一定の効果を持つことが推測される。オートエンコーダによって抽出した特徴量の1時刻分の同化実験では、コントロール実験と比較して、ほとんど変化がなかった。しかし、カルマンフィルタ (Kalman Filter, Kalman 1960) は誤差の確率分布がガウス分布であることを前提としており、この特徴量はガウス分布に類似していることから、長期間分の実験行うことで、降水領域の改善が期待できると考えられる。

機械学習モデルの列挙

D1-22

原 聡（大阪大学），石畠 正和（北海道大学），前原 貴憲（理研AIP）

あなたの機械学習モデル、信頼できますか？

- 回帰係数の値、おかしくないですか？
- 見つかった予測ルールに納得できますか？

- こんなことはありませんか？

回帰係数がおかしい。このモデル大丈夫？

エンジンの出力に寄与するはずのエンジン回転数の係数がゼロだぞ。。。

ルールに納得できない。このモデル大丈夫？

血圧が高いと病気になりやすいはずなのに、ルールに血圧が出てこない！

機械学習モデルの列挙

D1-22

原 聡（大阪大学），石畠 正和（北海道大学），前原 貴憲（理研AIP）

- **信頼できるモデル** って何？
  - 予測が当たればOK？ – No! それだけでは不十分
  - 納得できればOK？ – Yes! 納得できないものは使えない
- **信頼できるモデル** はどうしたら見つかる？

本研究の仮説

- **信頼できるモデル** は **良いモデル** の中にある
- **良いモデル** = 学習精度が（比較的）高い

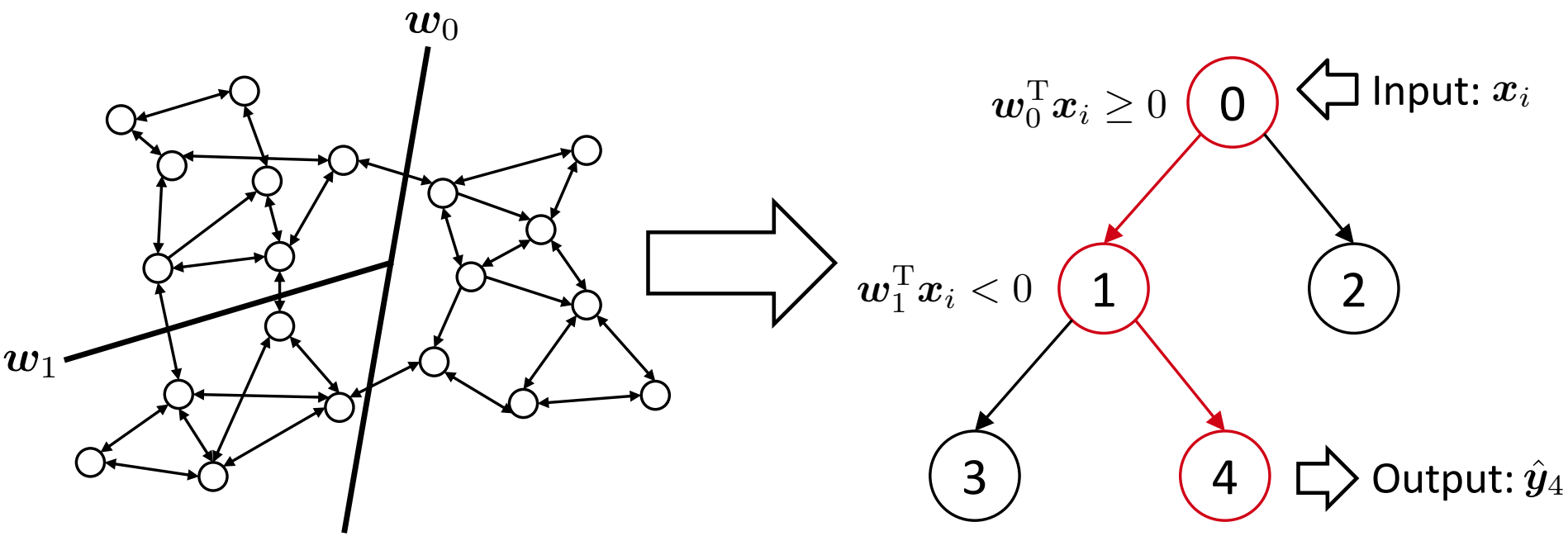
本研究の貢献

- **良いモデル** の列挙手法を提案
- 対象：線形モデル、ルールモデル

D1-23 k-最近傍グラフの分割による Extreme Multi-label分類器の学習

田頭 幸浩（ヤフー株式会社）

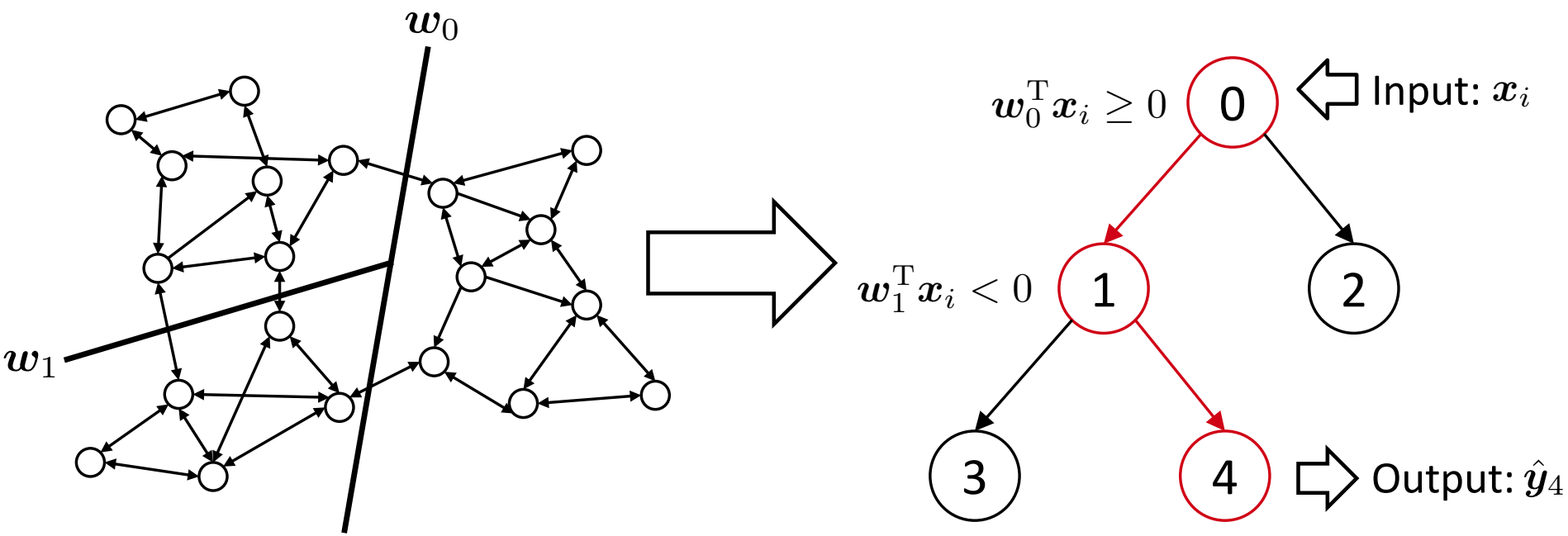
- ラベルの種類数が極端に多いマルチラベル分類問題
- ラベルベクトルのk-最近傍グラフの構造を保ったまま、特徴量空間で再帰的に分割することで、分類木を構築



D1-23 k-最近傍グラフの分割による Extreme Multi-label分類器の学習

田頭 幸浩（ヤフー株式会社）

- ラベルの種類数が極端に多いマルチラベル分類問題
- ラベルベクトルのk-最近傍グラフの構造を保ったまま、特徴量空間で再帰的に分割することで、分類木を構築

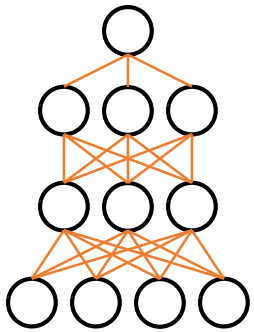


Deep reinforcement learning with relative entropy stochastic search (D1-24)

Voot Tangkaratt[1], Abbas Abdolmaleki[2], Masashi Sugiyama[1,3]

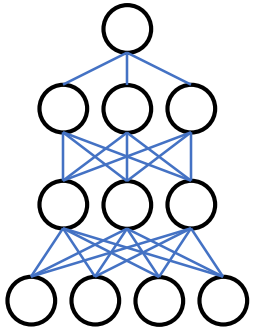
- **Actor-critic** (AC) methods are prominent deep RL methods.

Critic: $\widehat{Q}(\boldsymbol{s}, \boldsymbol{a})$



$\nabla_{\boldsymbol{\theta}} \mathcal{J}$

Actor: $\pi_{\boldsymbol{\theta}}(\boldsymbol{a}|\boldsymbol{s})$ or $\pi_{\boldsymbol{\theta}}(\boldsymbol{s})$



Update actor:

$$\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} + \alpha \nabla_{\boldsymbol{\theta}} \mathcal{J}$$

- Existing methods use **up to the 1st-order** information of critic
  - 0th-order (values): $\nabla_{\boldsymbol{\theta}} \mathcal{J} = \nabla_{\boldsymbol{\theta}} \log \pi_{\boldsymbol{\theta}}(\boldsymbol{a}|\boldsymbol{s}) \widehat{Q}(\boldsymbol{s}, \boldsymbol{a})$ (Sutton et al., 1999)
  - 1st-order (gradients): $\nabla_{\boldsymbol{\theta}} \mathcal{J} = \nabla_{\boldsymbol{\theta}} \pi_{\boldsymbol{\theta}}(\boldsymbol{s}) \nabla_{\boldsymbol{a}} \widehat{Q}(\boldsymbol{s}, \boldsymbol{a})$ (Silver et al., 2014)

- We propose **guide AC** (GAC) that uses **2nd-order** information (Hessians)

[1]RIKEN AIP（理化学研究所革新知能統合研究センター）
[2]The University of Aveiro, Portugal
[3]The University of Tokyo（東京大学）

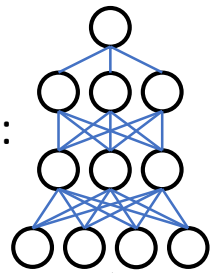
Guide actor-critic (GAC)

- GAC learning steps:
  1. Learn a guide actor $\pi_+(\boldsymbol{a}|\boldsymbol{s})$
  2. Update the actor $\pi_{\boldsymbol{\theta}}(\boldsymbol{a}|\boldsymbol{s})$ to fit $\pi_+(\boldsymbol{a}|\boldsymbol{s})$
  3. Repeat until convergence

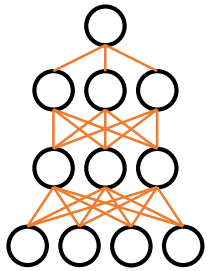
$$\pi_{\boldsymbol{\theta}}(\boldsymbol{a}|\boldsymbol{s}) = \mathcal{N}(\boldsymbol{a}|\boldsymbol{\phi}_{\boldsymbol{\theta}}(\boldsymbol{s}), \boldsymbol{\Sigma})$$

$$\widehat{Q}(\boldsymbol{s}, \boldsymbol{a})$$

Actor:



Critic:



$$\boldsymbol{g} = \nabla_{\boldsymbol{a}} \widehat{Q}(\boldsymbol{s}, \boldsymbol{a})$$

$$\boldsymbol{H} = \nabla^2_{\boldsymbol{a}} \widehat{Q}(\boldsymbol{s}, \boldsymbol{a})$$

1. Learn guide actor:

$$\pi_+(\boldsymbol{a}|\boldsymbol{s}) = \mathcal{N}(\boldsymbol{a}|\boldsymbol{\phi}_+(\boldsymbol{s}), \boldsymbol{\Sigma}_+)$$

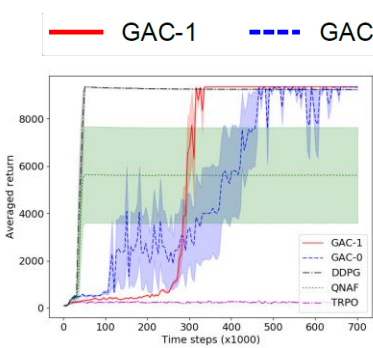
$$\boldsymbol{\phi}_+(\boldsymbol{s}) = \boldsymbol{\phi}(\boldsymbol{s}) + (\eta\boldsymbol{\Sigma} - \boldsymbol{H})^{-1}\boldsymbol{g}$$

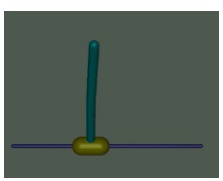
$$\boldsymbol{\Sigma}_+ = \beta(\eta\boldsymbol{\Sigma}^{-1} + \boldsymbol{H})^{-1}$$

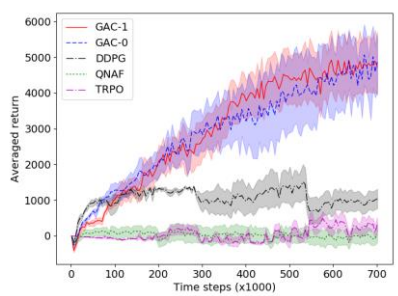
2. Update actor:

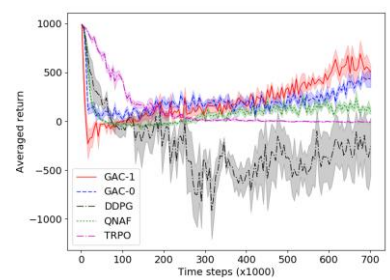
$$\min_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}) := \|\boldsymbol{\phi}_{\boldsymbol{\theta}} - \boldsymbol{\phi}_+\|^2$$

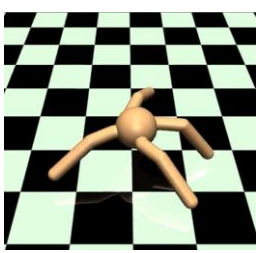
$$\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} - \alpha \nabla_{\boldsymbol{\theta}} \mathcal{L}$$







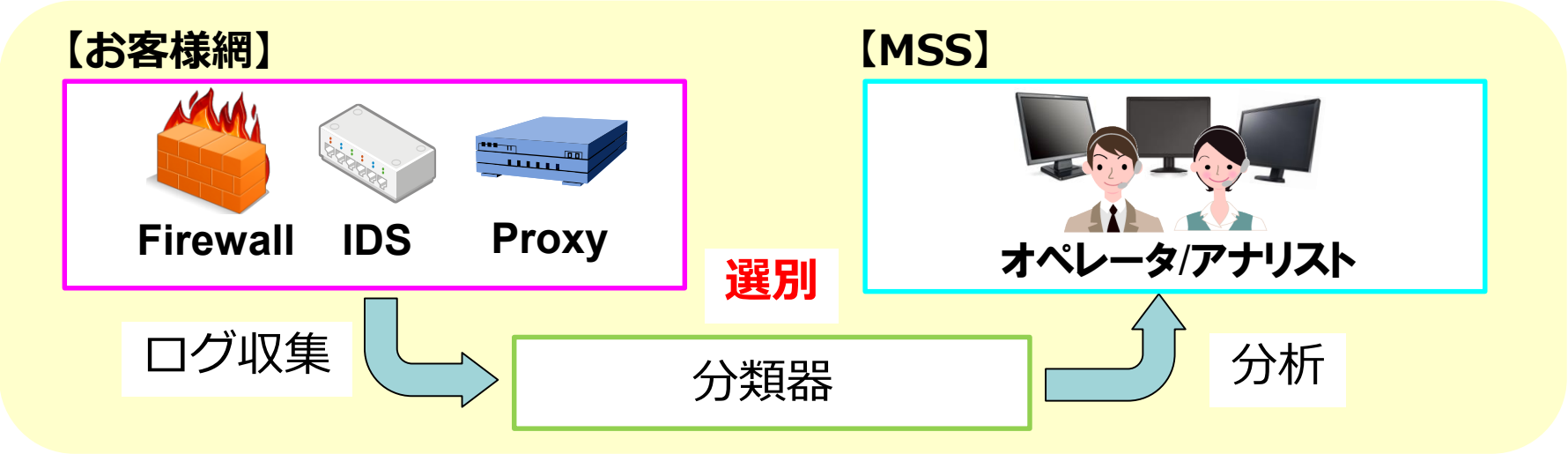




D1-25:HTTP通信を用いた半教師あり学習による マルウェア感染端末の分類器について

西山 泰史, 熊谷 充敏, 神谷 和憲, 谷川 真樹　（NTT）

- **背景**：マルウェア感染の有無を判定する分類器が必要
  ⇒機械学習で作成することで運用負担を軽減
- **提案**：半教師あり学習を用いた分類器作成法
  -結果の解釈性があり、検知理由が分析可能な手法
- **結果**：検知できるマルウェアが増加/分類性能向上

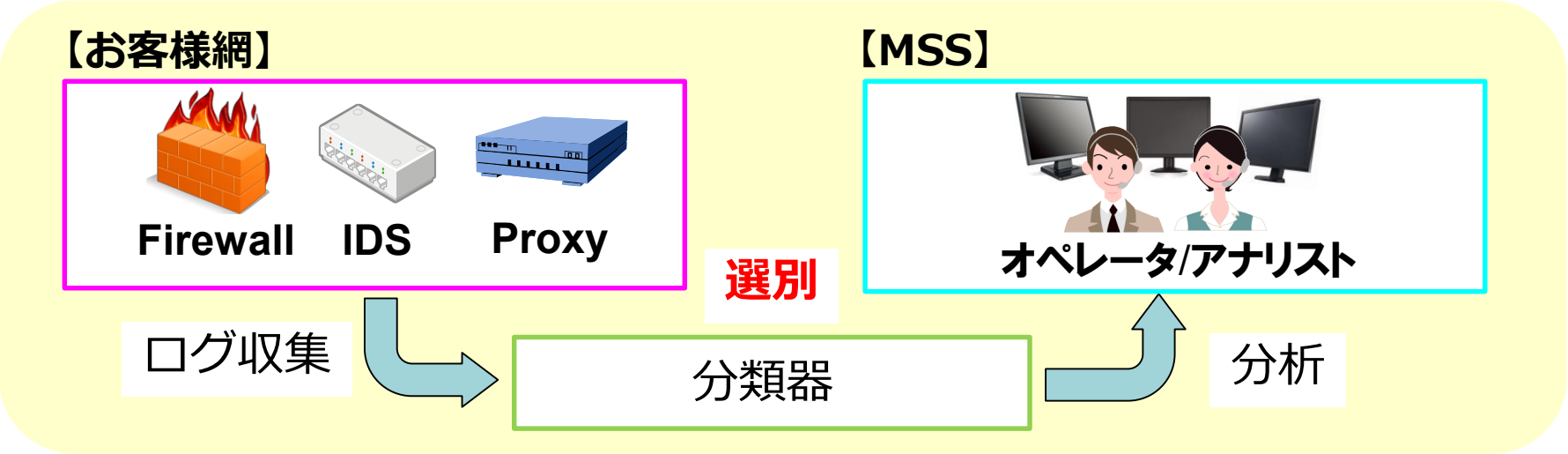


Copyright©2017 NTT corp. All Rights Reserved.

D1-25:HTTP通信を用いた半教師あり学習による マルウェア感染端末の分類器について

西山 泰史, 熊谷 充敏, 神谷 和憲, 谷川 真樹　（NTT）

- **背景**：マルウェア感染の有無を判定する分類器が必要
  ⇒機械学習で作成することで運用負担を軽減
- **提案**：半教師あり学習を用いた分類器作成法
  -結果の解釈性があり、検知理由が分析可能な手法
- **結果**：検知できるマルウェアが増加/分類性能向上



Copyright©2017 NTT corp. All Rights Reserved.

D1-26：生成要約のための隣接畳み込みニューラルネットワーク

近藤 雅芳, 松本 裕治 (奈良先端科学技術大学院大学, 理研AIP-知識獲得チーム)

＜課題＞ 複数文章(長文)を入力とする生成要約は、RNN型encoderでは **訓練時の勾配爆発・消失** と **逐次エンコードの処理時間増大** が課題.

【概要】：長文の生成要約において、RNN型encoderと代替え可能な **新しいCNN型encoderの構築とその精度検証.**

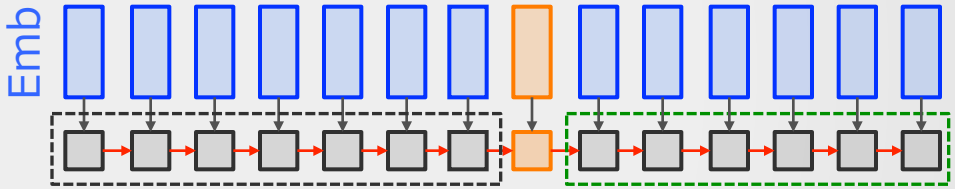
提案手法：

隣接畳み込みニューラルネット
(Adjoined Convolutional Neural Network)

- エンコード対象の系列要素の両隣接要素を畳み込む & Gate関数化
- 系列データの並列エンコード
- 長期系列に対する勾配伝搬の回避

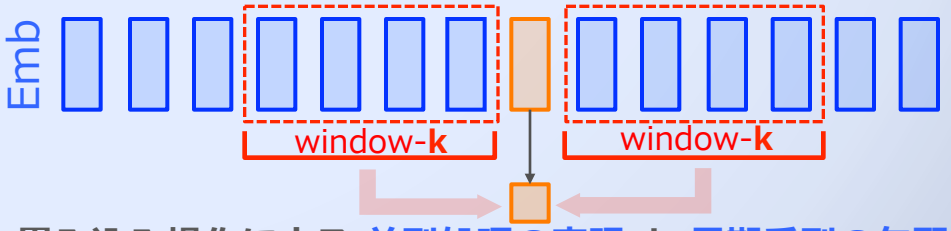
【精度検証・評価】
- **データセット**
  > **CNN/DailyMail** データセット
- **評価指標**
  > **ROUGE-スコア**(R-1/R-2/R-L)

【従来の一般的手法：RNN型 -Encoder】



逐次エンコードによる処理時間増加 と **勾配爆発・消失**

【提案手法：隣接畳み込みニューラルネット】



畳み込み操作による 並列処理の実現 と 長期系列の勾配伝搬の回避

D1-26：生成要約のための隣接畳み込みニューラルネットワーク

近藤 雅芳, 松本 裕治 (奈良先端科学技術大学院大学, 理研AIP-知識獲得チーム)

＜課題＞ 複数文章(長文)を入力とする生成要約は、RNN型encoderでは **訓練時の勾配爆発・消失** と **逐次エンコードの処理時間増大** が課題.

【概要】：長文の生成要約において、RNN型encoderと代替え可能な **新しいCNN型encoderの構築とその精度検証.**

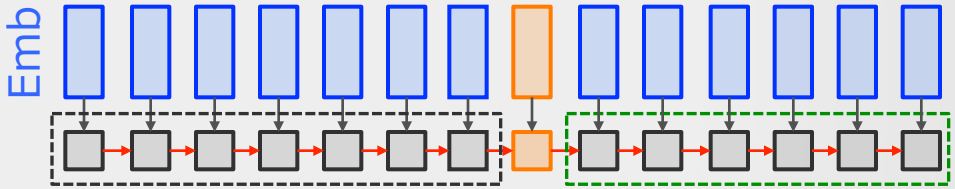
提案手法：

隣接畳み込みニューラルネット
(Adjoined Convolutional Neural Network)

- エンコード対象の系列要素の両隣接要素を畳み込む & Gate関数化
- 系列データの並列エンコード
- 長期系列に対する勾配伝搬の回避

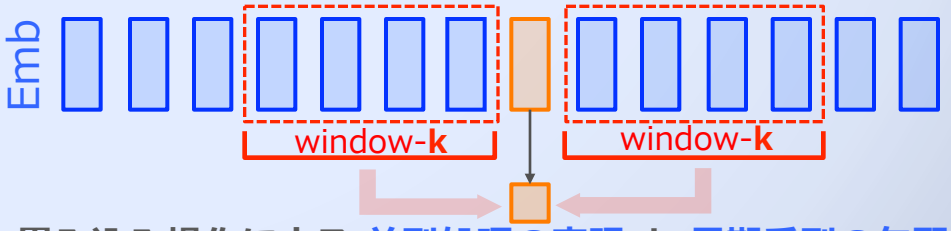
【精度検証・評価】
- **データセット**
  > **CNN/DailyMail** データセット
- **評価指標**
  > **ROUGE-スコア**(R-1/R-2/R-L)

【従来の一般的手法：RNN型 -Encoder】



逐次エンコードによる処理時間増加 と **勾配爆発・消失**

【提案手法：隣接畳み込みニューラルネット】



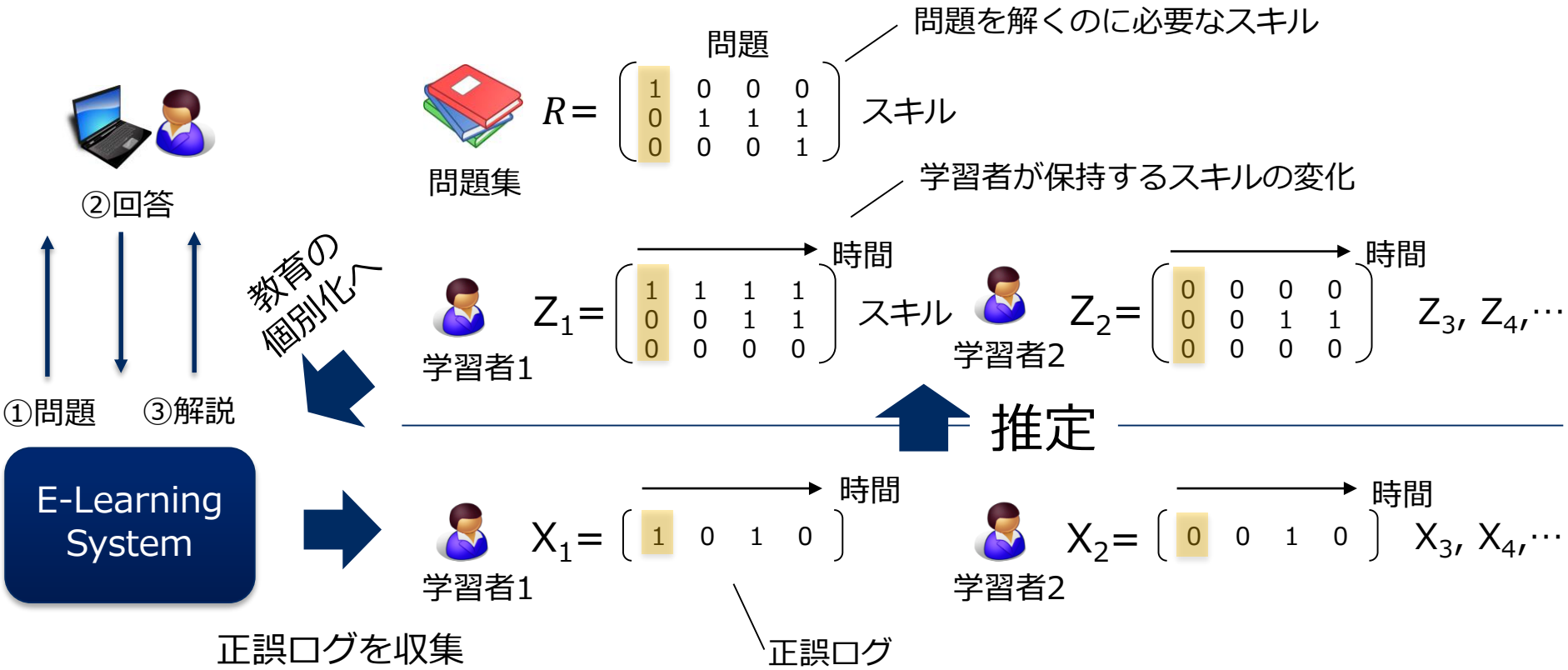
畳み込み操作による 並列処理の実現 と 長期系列の勾配伝搬の回避

D1-27: 教育ログデータからの解釈性を重視した潜在スキル推定

玉野浩嗣、丸山晃一（NECソリューションイノベータ株式会社）

背景：学習者が問題に取り組んだ正誤ログから、学習者が保持する潜在スキルの変化と、問題を解くのに必要な潜在スキルを推定する問題

目的：推定結果を人間が解釈しやすいバイナリ値で推定する方法を提案



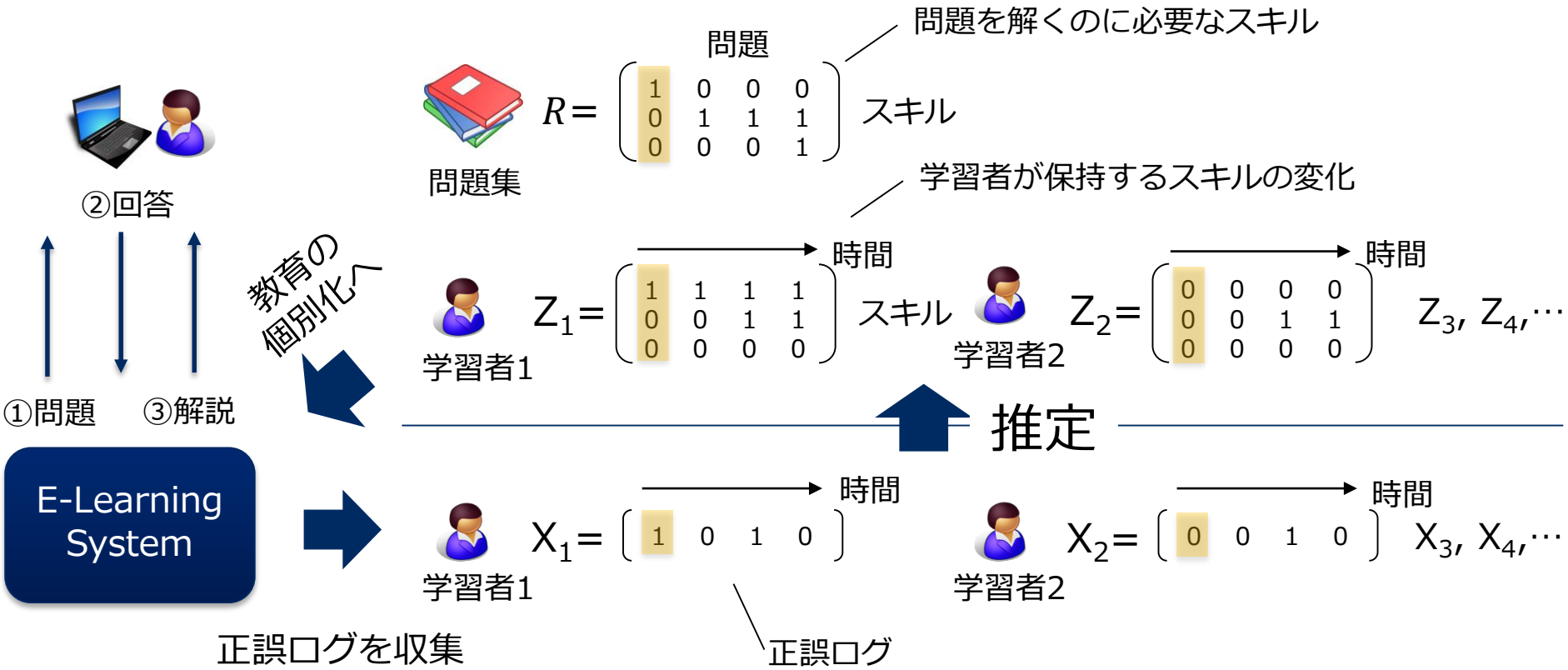
© NEC Solution Innovators, Ltd. 2017

D1-27: 教育ログデータからの解釈性を重視した潜在スキル推定

玉野浩嗣、丸山晃一（NECソリューションイノベータ株式会社）

背景：学習者が問題に取り組んだ正誤ログから、学習者が保持する潜在スキルの変化と、問題を解くのに必要な潜在スキルを推定する問題

目的：推定結果を人間が解釈しやすいバイナリ値で推定する方法を提案



© NEC Solution Innovators, Ltd. 2017

D1-28

Bayesian NetworkとDeep Learningによる品質予測方式の開発

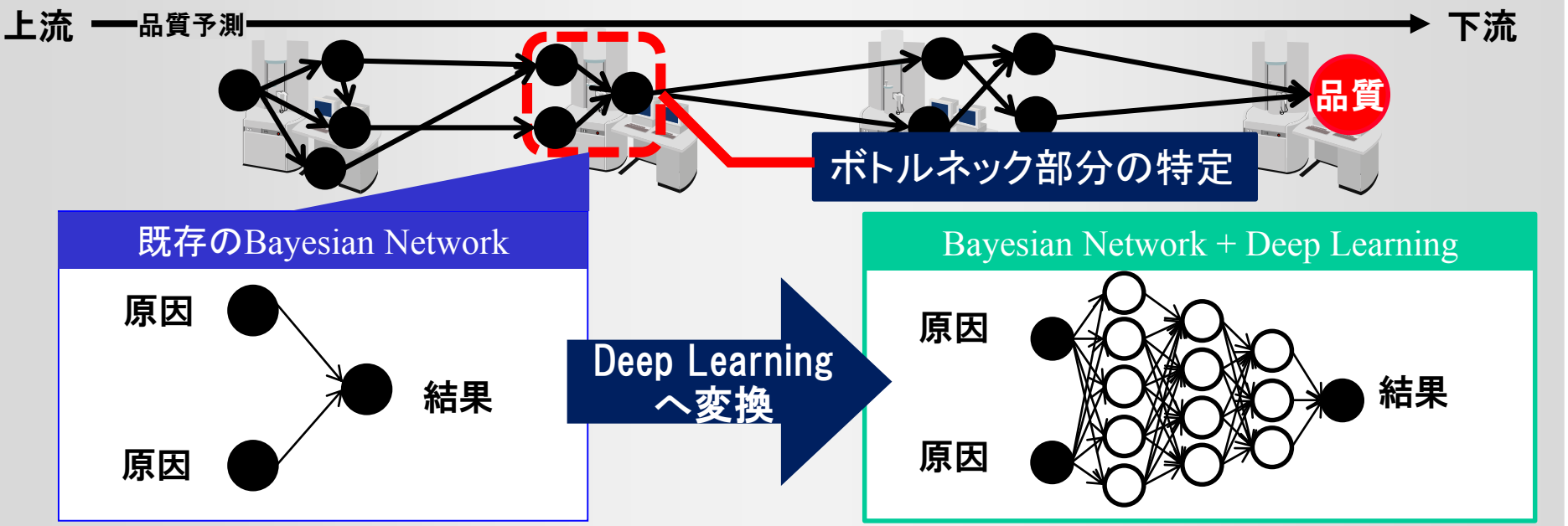
HITACHI
Inspire the Next

株式会社 日立製作所
研究開発グループ　生産システム研究部
堀脇一樹, 今沢 慶

研究対象	大規模製造工程におけるBayesian Networkによる品質予測・制御
問題点	大規模製造工程における品質予測時の誤差累積

問題に対するアプローチと提案方式

品質予測時に誤差累積を生み出す
ボトルネック部分を特定し、Deep Learningへ変換



© Hitachi, Ltd. 2017. All rights reserved.

D1-28

Bayesian NetworkとDeep Learningによる品質予測方式の開発

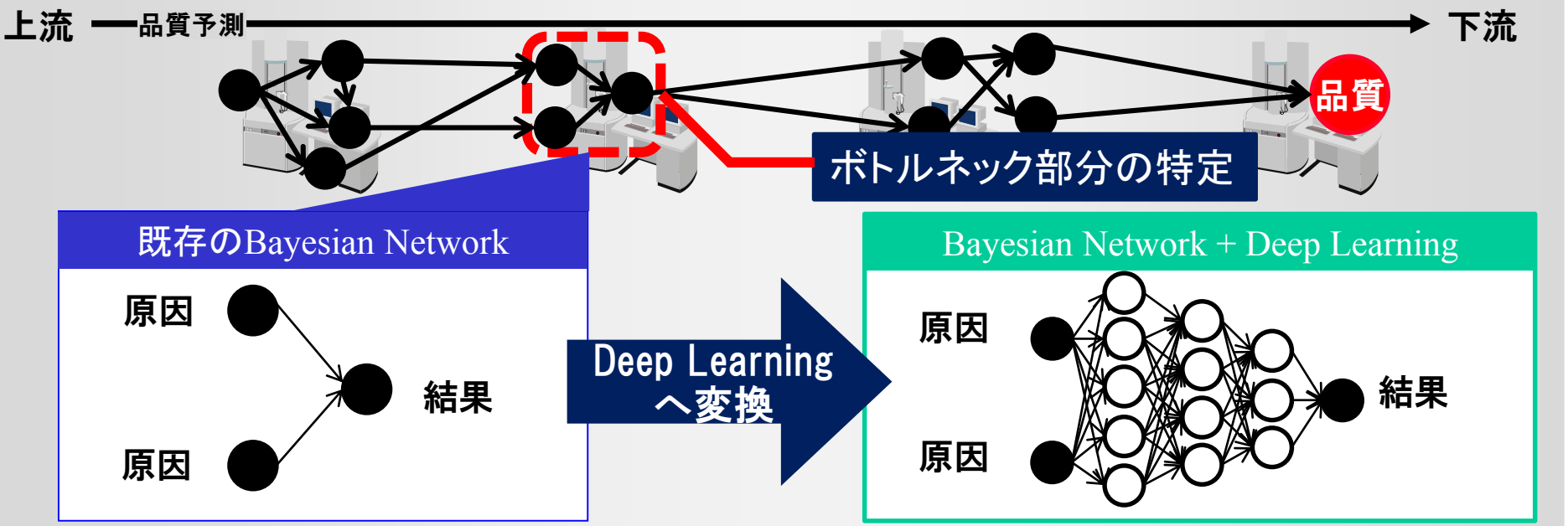
HITACHI
Inspire the Next

株式会社 日立製作所
研究開発グループ　生産システム研究部
堀脇一樹, 今沢 慶

研究対象	大規模製造工程におけるBayesian Networkによる品質予測・制御
問題点	大規模製造工程における品質予測時の誤差累積

問題に対するアプローチと提案方式

品質予測時に誤差累積を生み出す
ボトルネック部分を特定し、Deep Learningへ変換



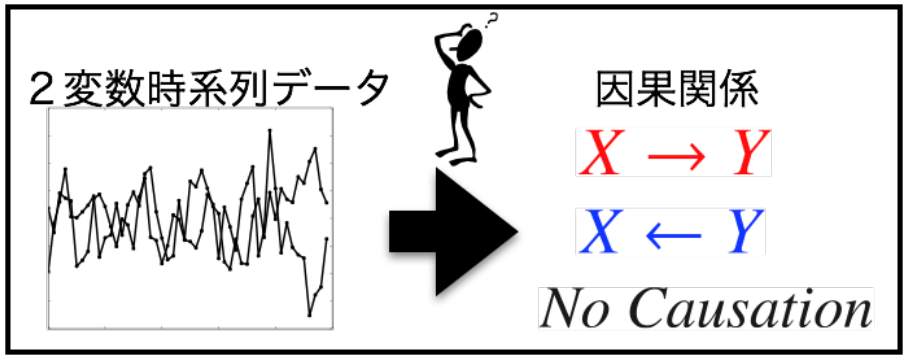
© Hitachi, Ltd. 2017. All rights reserved.

D1
29

教師あり学習に基づく時系列の因果推論

近原 鷹一　　藤野 昭典
NTTコミュニケーション科学基礎研究所

タスク： 時系列データから因果関係を推定



例

X: R&D expenditure
Y: total sales

$X \rightarrow Y$

アプローチ：分類問題として解く

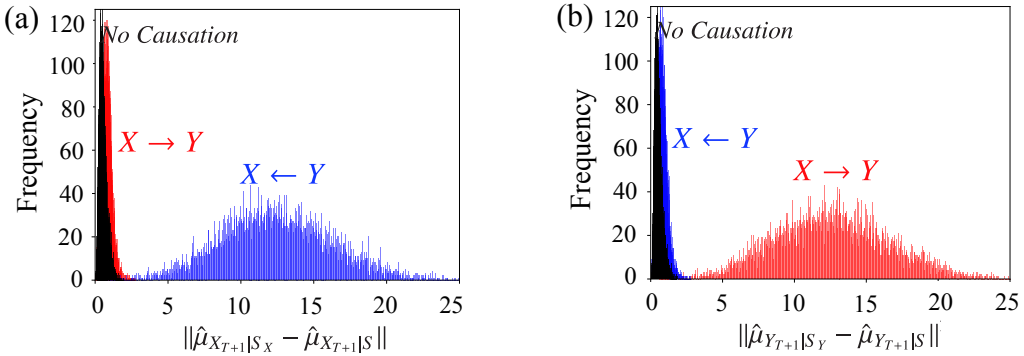
時系列データに対し
因果関係を表すクラスラベルを割り当てる

D1
29

教師あり学習に基づく時系列の因果推論

近原 鷹一　　藤野 昭典
NTTコミュニケーション科学基礎研究所

貢献：<u>因果関係によって異なる特徴ベクトルが得られる</u>
特徴量表現を提案



結果： 人工データ・実データともに既存手法より高精度

Key Words：
1. Granger Causality
2. Kernel Embeddings
3. Time Series Analysis

D1-30: ロバスト制御方策と探索木を用いた近似ベイジアン強化学習アプローチ

菱沼徹，泉田啓 （京都大学）

目的

- ロボットの自律制御に強化学習を応用したい.
- 実ロボットの動作コストを抑えつつ，現実的な計算量で実装したい.

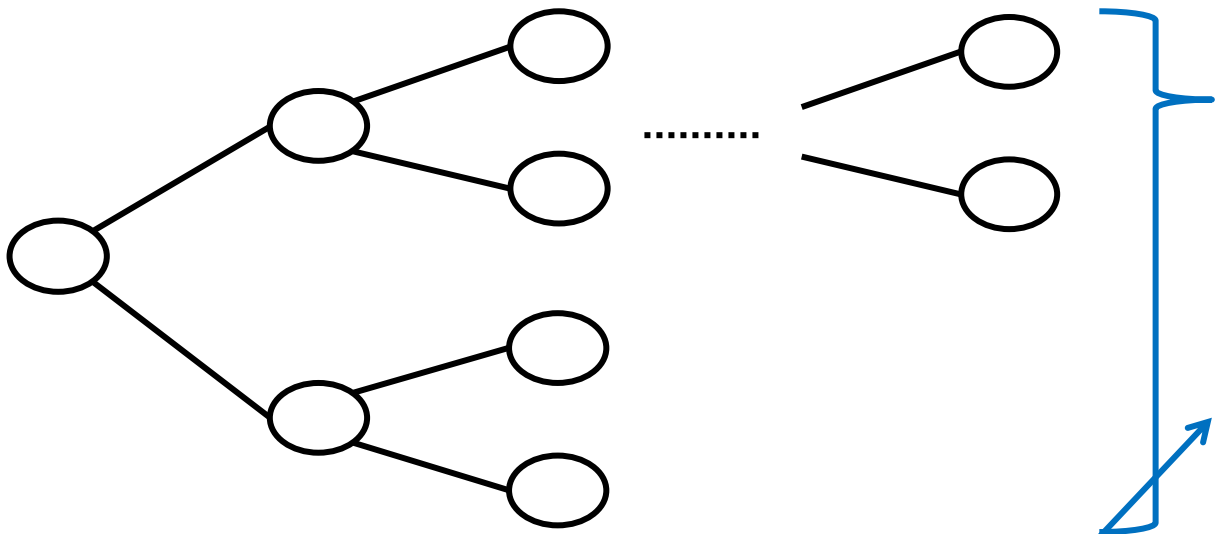
アプローチ

- モデルベース・ベイジアン強化学習（MB-BRL）問題.
  - 事前知識として運動方程式モデルを仮定，そのパラメータを未知とする.
- より計算量的に取り扱いやすい，近似MB-BRL問題.
  - 前半：有限深さ探索木問題，後半：ロバスト制御問題.
- この近似MB-BRL問題を解くための，シミュレーション解法
  - モンテカルロ木探索，ロバスト強化学習，近似ベイズ推定，の組合せ.

近似問題とアルゴリズムの概要

Finite horizon 探索木問題
履歴依存方策で，オンライン情報をフル活用

ロバスト制御問題
定常方策で，現在の状態のみを考慮



ロバスト強化学習
各(s,b)に対してある程度の性能を持つ定常方策をシミュレーション学習

近似ベイズ推定
履歴に対する信念更新を，MCTSシミュレーションのサンプルを用いて近似

(s,b)
（状態，信念）

u
行動

(s,b)

(s,b)

timestep

モンテカルロ木探索（MCTS）
finite horizon問題のシミュレーション解法

この近似問題における価値関数を，運動方程式シミュレーション上で学習し，その結果を実環境上で用いる．

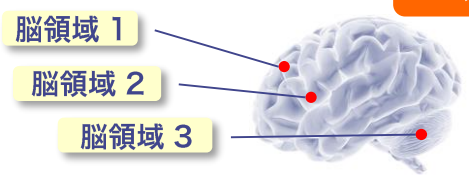
D1-31

Sparse Regression for Network Graphs and Its Application to Gene Networks of The Brain

Hideko Kawakubo[1], Yusuke Matsui[1], Teppei Shimamura[1], [1]Nagoya University

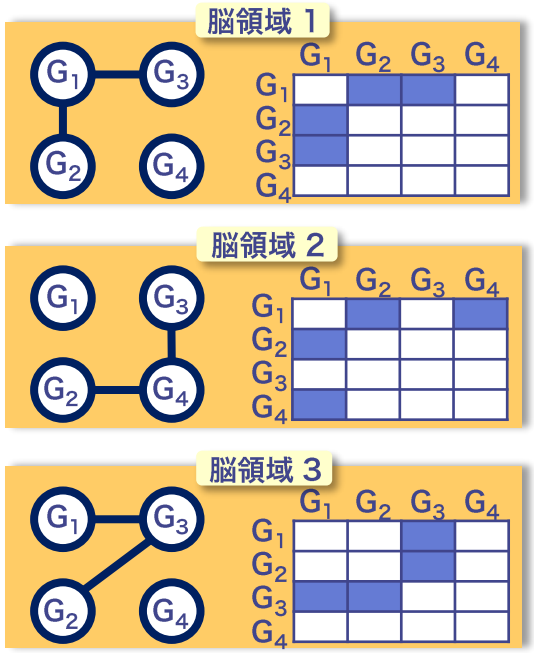
脳部位特異的遺伝子ネットワークと疾患との関連性を調べたい

入力



脳部位特異的ネットワーク

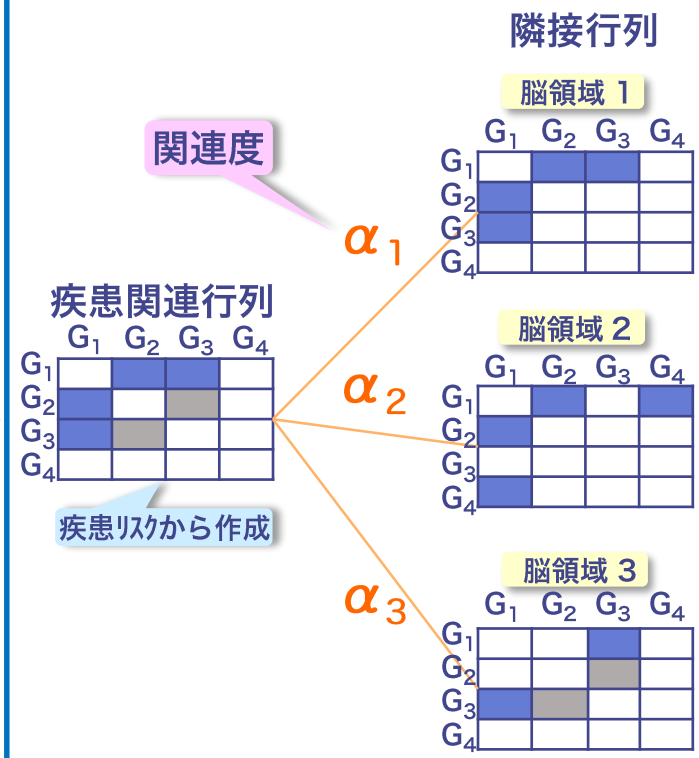
グラフ　　隣接行列



疾患リスク

Gene ID	リスク
G_1	y_1
G_2	y_2
G_3	y_3
G_4	y_4

GOSPEL（提案法）



脳領域間の相関を考慮し, **疾患関連行列を隣接行列の線形和で表現**

シミュレーションの結果、提案法は従来法よりも約15%精度が向上した

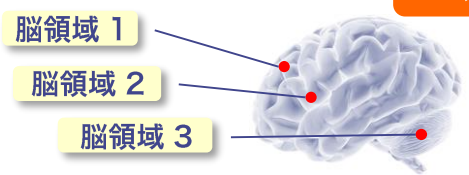
D1-31

Sparse Regression for Network Graphs and Its Application to Gene Networks of The Brain

Hideko Kawakubo[1], Yusuke Matsui[1], Teppei Shimamura[1], [1]Nagoya University

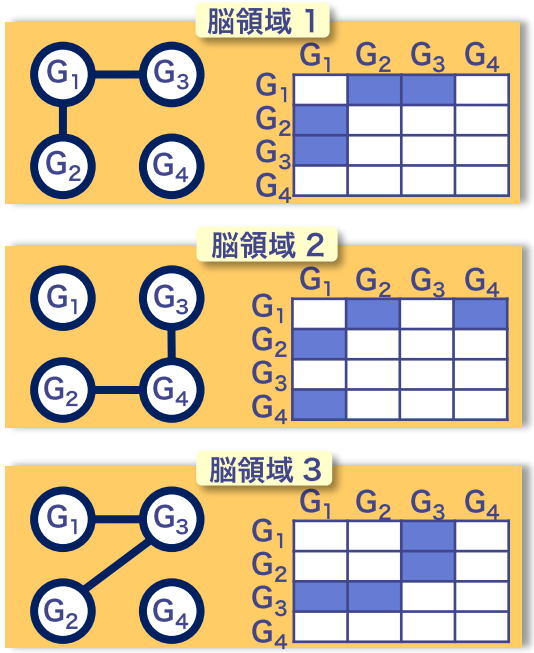
脳部位特異的遺伝子ネットワークと疾患との関連性を調べたい

入力



脳部位特異的ネットワーク

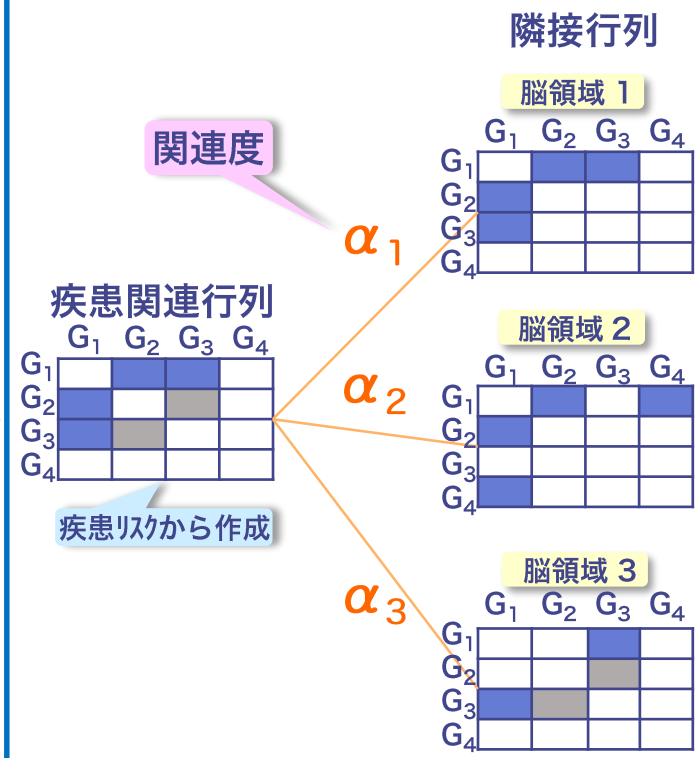
グラフ　　隣接行列



疾患リスク

Gene ID	リスク
G_1	y_1
G_2	y_2
G_3	y_3
G_4	y_4

GOSPEL（提案法）



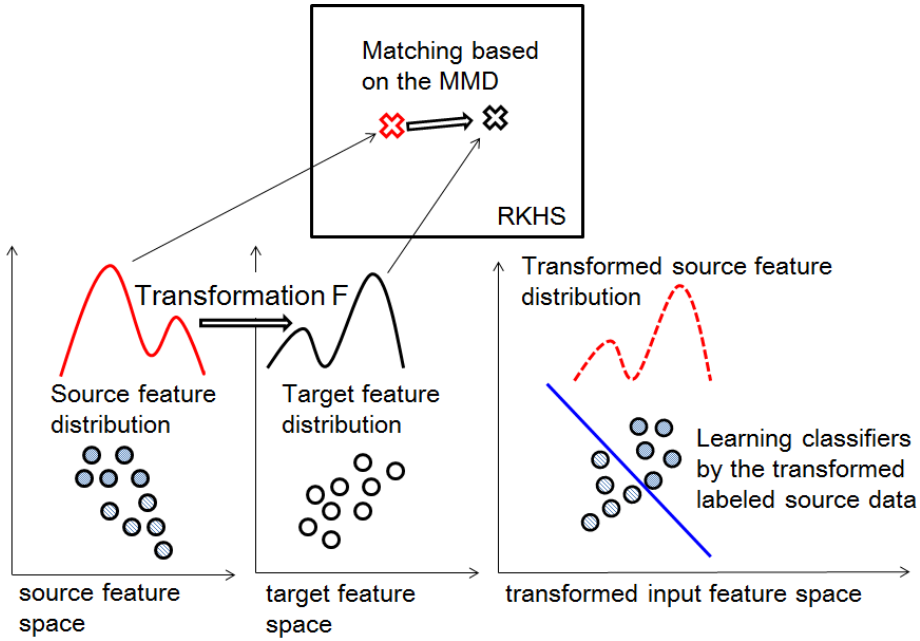
脳領域間の相関を考慮し, **疾患関連行列を隣接行列の線形和で表現**

シミュレーションの結果、提案法は従来法よりも約15%精度が向上した

D1-32: Maximum mean discrepancyに基づく分布マッチングを用いた教師なしドメイン適応

熊谷充敏, 岩田具治（NTT）

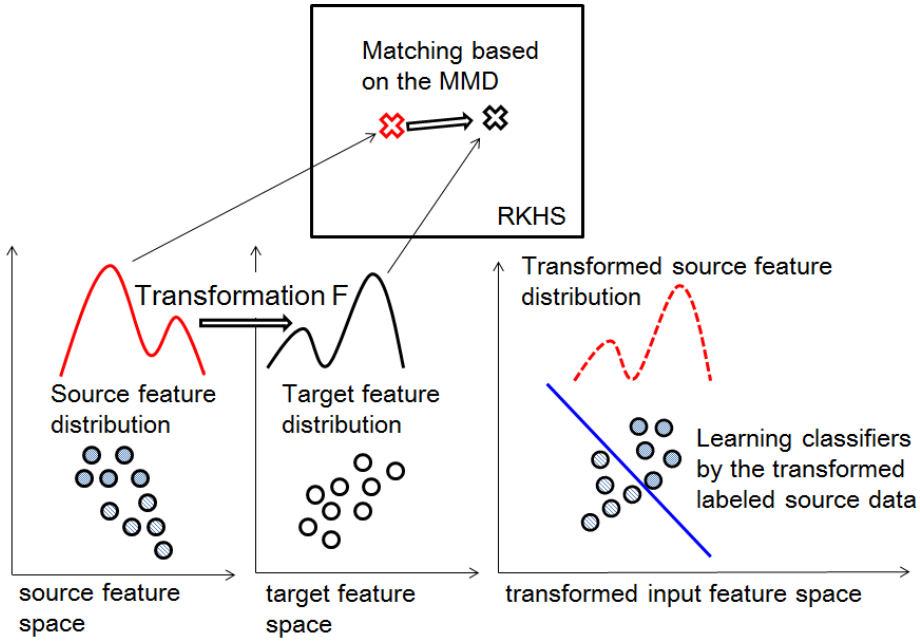
- 新たな教師なしドメイン適応法を提案
- Maximum mean discrepancyに基づき，元ドメインと目標ドメインの特徴分布が同じになるよう，元ドメインから目標ドメインへの変換則を学習
- 実データを用いた実験で有効性を確認



D1-32: Maximum mean discrepancyに基づく分布マッチングを用いた教師なしドメイン適応

熊谷充敏, 岩田具治（NTT）

- 新たな教師なしドメイン適応法を提案
- Maximum mean discrepancyに基づき，元ドメインと目標ドメインの特徴分布が同じになるよう，元ドメインから目標ドメインへの変換則を学習
- 実データを用いた実験で有効性を確認



D1-33　LogDet正則化を用いた二値行列補完

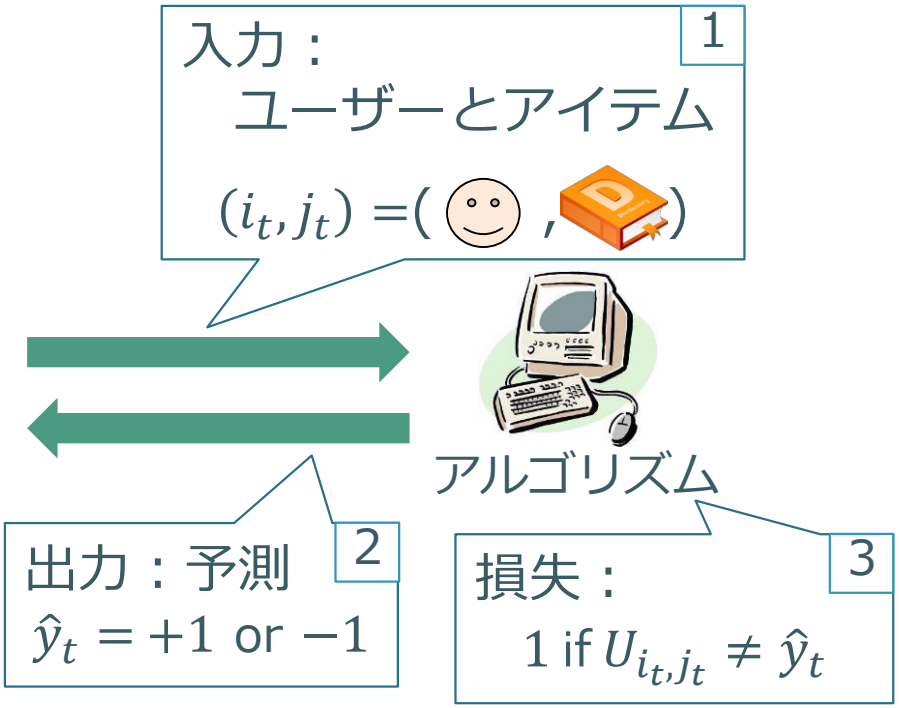
森富 賢一郎・畑埜 晃平・瀧本 英二（九州大学）

二値行列補完

For $t = 1, \ldots, T$

$m \times n$ 二値行列 U

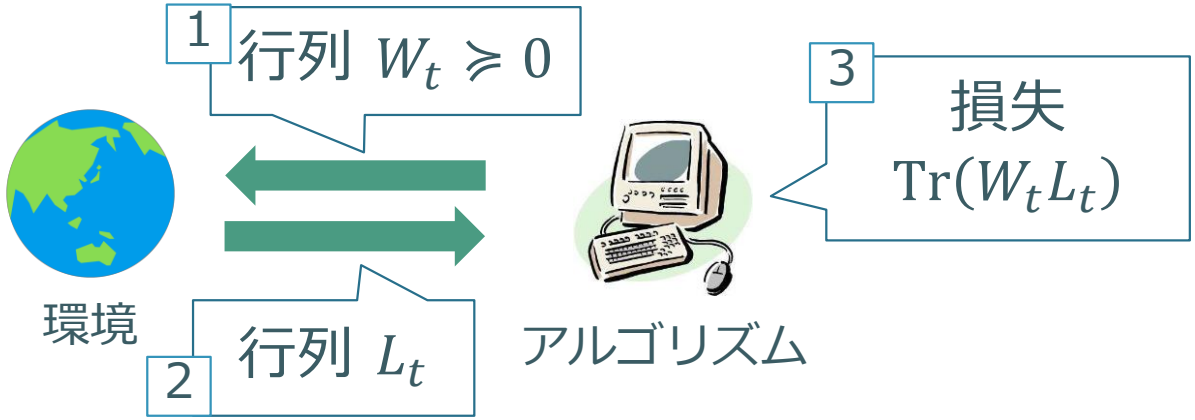
U	1	2	3	4	5
	?	+	+	?	+
	−	?	−	?	−
	?	+	−	−	?



提案手法

半正定値対称行列のオンライン予測問題に帰着

For $t = 1, \ldots, T$



アルゴリズム　Follow the Regularized Leader

$\mathcal{K}$ は半正定値対称行列の有界凸集合

正則化項 R と学習率 η でアルゴリズムが決まる

$$W_{t+1} = \mathrm{argmin}_{W \in \mathcal{K}} \left(R(W) + \eta \sum_{s=1}^{t} \mathrm{Tr}(W L_s) \right)$$

	正則化項
Herbstar et al., 2016	負のエントロピー $R(W) = \mathrm{Tr}(W \log W - W)$
本研究	Log-Determinant $R(W) = -\log \det(W + \epsilon \mathrm{E})$

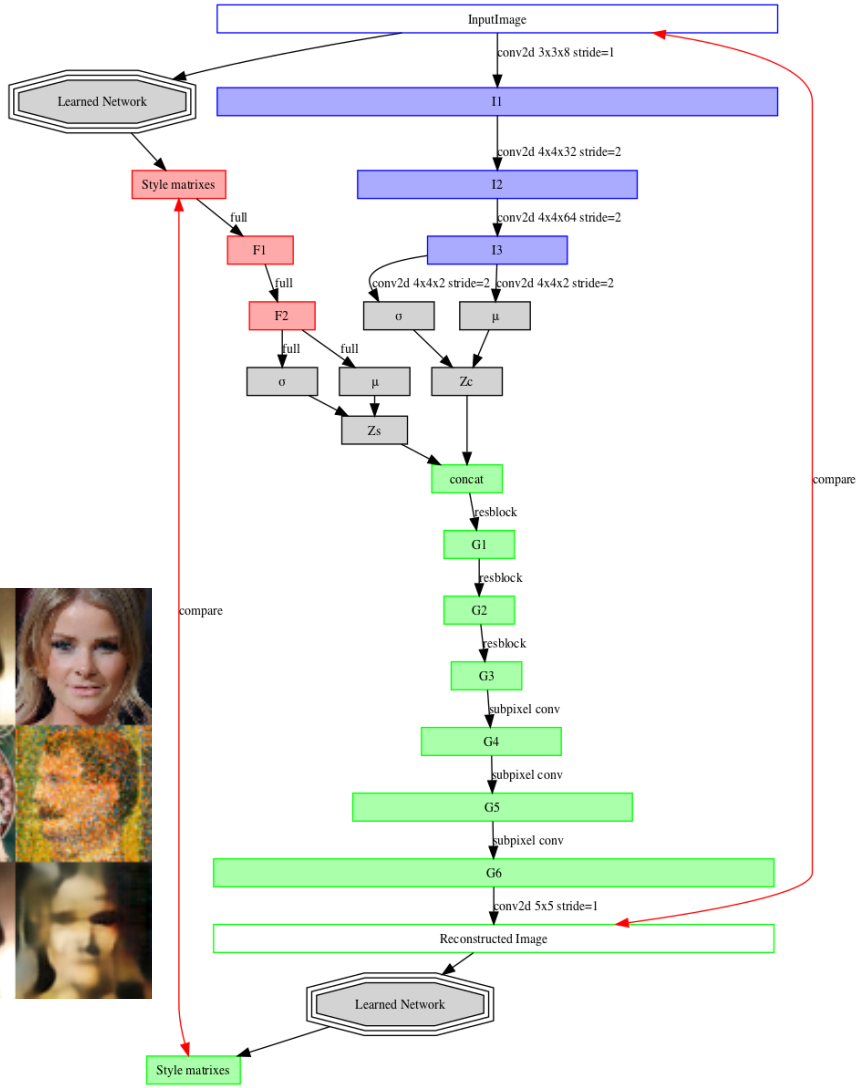
主結果

- 誤り回数の上界を導く
  - 既存手法より $\log(m+n)$ 倍改善，下界と一致し最適
- 統計学習へ応用，汎化誤差を導出
  - 最適なカーネルを知らずとも，それと同等の効果

D1-34: VAEを用いた画像スタイル変換に向けて

中田秀基、麻生英樹 （産業技術総合研究所 人工知能研究センター）

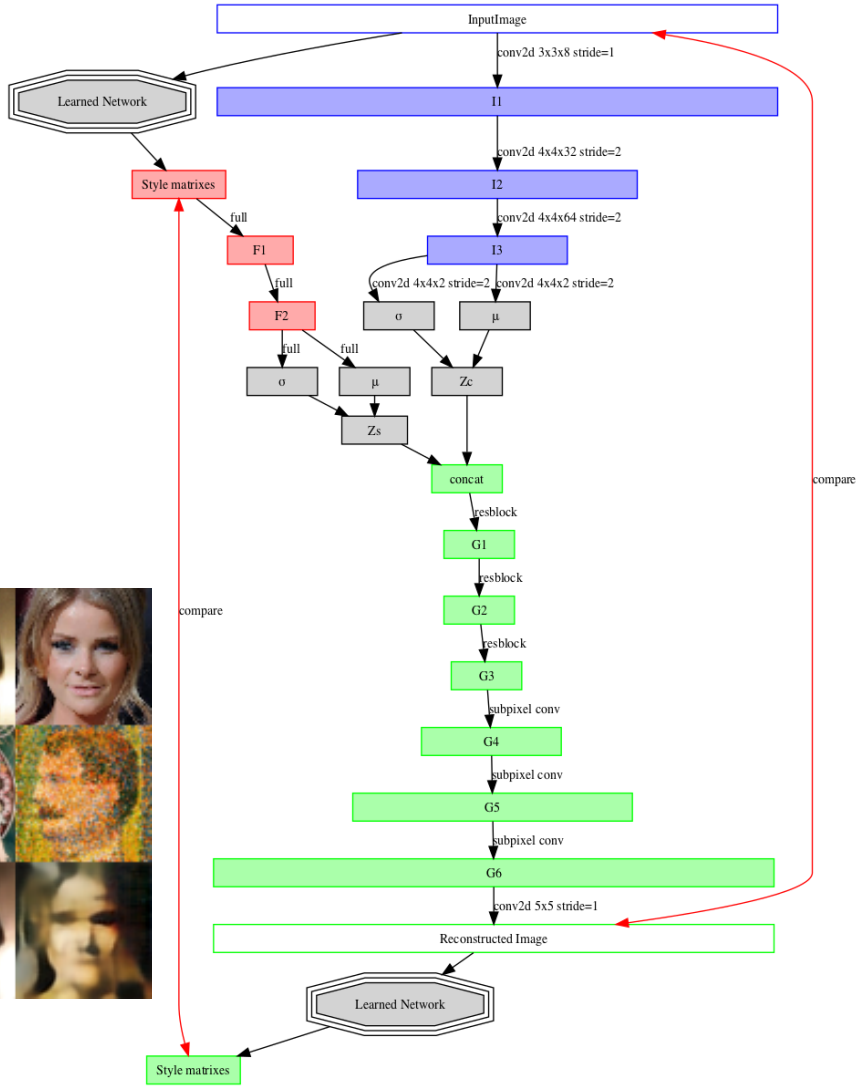
- 一つのネットワークで任意のスタイルへの変換を実現することを目指す
- スタイルとコンテンツを分離して隠れ変数に学習
- コンテンツ画像から作った隠れ変数のスタイル部分を、スタイル画像から作った隠れ変数部分で置き換えることで、スタイル変換を実現



D1-34: VAEを用いた画像スタイル変換に向けて

中田秀基、麻生英樹 （産業技術総合研究所 人工知能研究センター）

- 一つのネットワークで任意のスタイルへの変換を実現することを目指す
- スタイルとコンテンツを分離して隠れ変数に学習
- コンテンツ画像から作った隠れ変数のスタイル部分を、スタイル画像から作った隠れ変数部分で置き換えることで、スタイル変換を実現



D1-35: 人ロボット協調作業ラインにおける ロボット異常検知技術の開発

HITACHI
Inspire the Next

- 製造業では，品質と生産性の両立のため，高難度作業を得意とする人と，繰り返し作業を得意とするロボットによる協調作業が求められている．
- ロボットは，人の作業状況に応じて，動作を変更するため，正常な動作パターンが，毎回異なる．
- そのため，過去の動作パターンとの比較によるロボット異常を検出することが困難である．
- 本発表では，**人と協調作業を行うロボットの手の軌跡データに対する異常検知方式**を発表する．

【問題設定】
正常な動作パターンが毎回異なる
ロボットの手の軌跡データに対する
異常検知方式の開発

■ロボットの動作
左手：人に部品を渡す
右手：人を避けながら作業を遂行する
→ 正常な動作パターンが毎回異なる

2017年11月9日（木）
（株）日立製作所　CTI 生産イノベーションセンタ
生産システム研究部生産S2ユニット研究員
兼 インダストリソリューションデザイン室員
今沢　慶

© Hitachi, Ltd. 2017. All rights reserved.

D1-35: 人ロボット協調作業ラインにおける ロボット異常検知技術の開発

HITACHI
Inspire the Next

- 製造業では，品質と生産性の両立のため，高難度作業を得意とする人と，繰り返し作業を得意とするロボットによる協調作業が求められている．
- ロボットは，人の作業状況に応じて，動作を変更するため，正常な動作パターンが，毎回異なる．
- そのため，過去の動作パターンとの比較によるロボット異常を検出することが困難である．
- 本発表では，**人と協調作業を行うロボットの手の軌跡データに対する異常検知方式**を発表する．

【問題設定】
正常な動作パターンが毎回異なる
ロボットの手の軌跡データに対する
異常検知方式の開発

■ロボットの動作
左手：人に部品を渡す
右手：人を避けながら作業を遂行する
→ 正常な動作パターンが毎回異なる

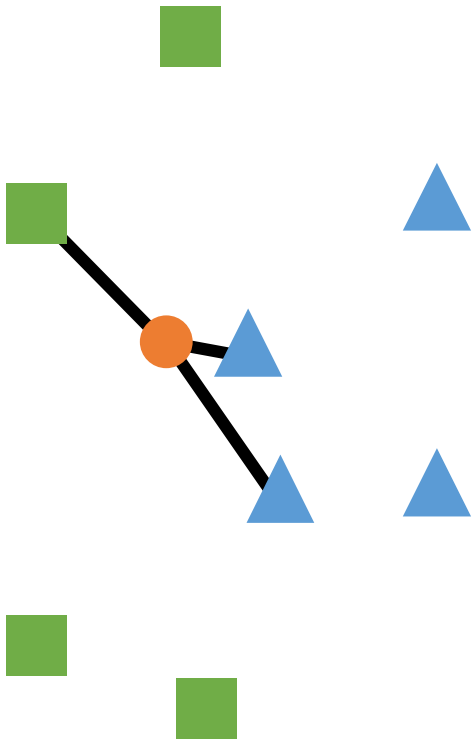
2017年11月9日（木）
（株）日立製作所　CTI 生産イノベーションセンタ
生産システム研究部生産S2ユニット研究員
兼 インダストリソリューションデザイン室員
今沢　慶

© Hitachi, Ltd. 2017. All rights reserved.

D1-36 南直人，延原肇（筑波大学）

単体複体の増大列を利用したマルチスケールk近傍法と多クラス点群データの幾何学的構造分析への応用

従来法（k近傍法）

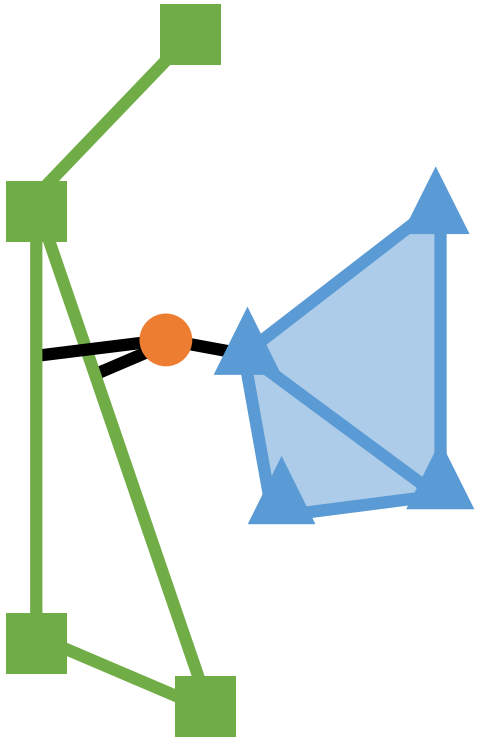


3NN(●) … [▲, ■, ▲]

CLASS(●) = ▲

提案法

（k近傍法の教師データに単体を追加）



3NN(●) … [▬, ▬, ▲]

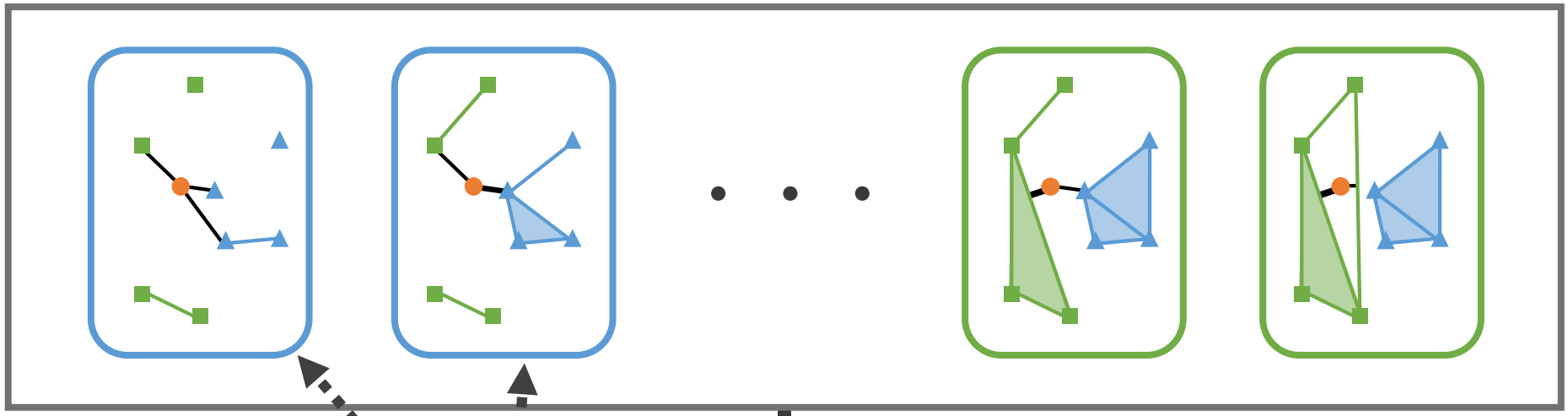
CLASS(●) = ■

D1-36 南直人，延原肇（筑波大学）

単体複体の増大列を利用したマルチスケールk近傍法と多クラス点群データの幾何学的構造分析への応用

さまざまな**スケール**で
データを捉えることができる

弱識別器の集合



スケール**小**

スケール**大**

Voting

識別

それぞれが
単体複体を用いたk近傍法による
識別器となっている

ポスターブース D1-36へ！

Robust and Efficient Off-Policy Policy Evaluation via Enhanced Action-Gap

Tadashi Kozuno (OIST), Eiji Uchibe (ATR & OIST), Kenji Doya (OIST) ID: D1-37

Problem

We rephrase policy evaluation problem as:

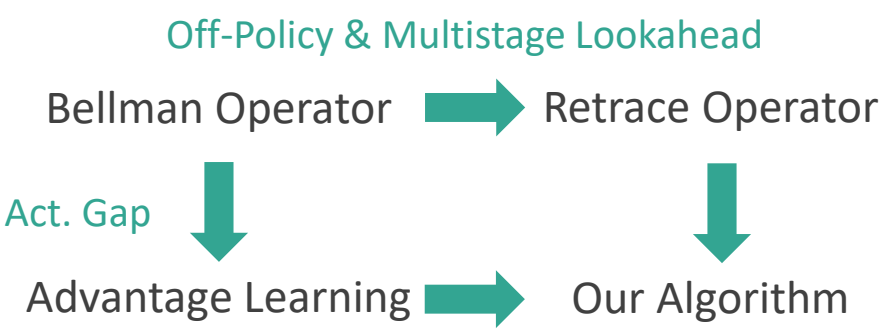
Given a policy π, order actions according to $Q^{\pi}(s,a) \coloneqq \mathbb{E}^{\pi}[\sum_{t\geq 0}\gamma^t R_t \mid s,a]$.

Contributions

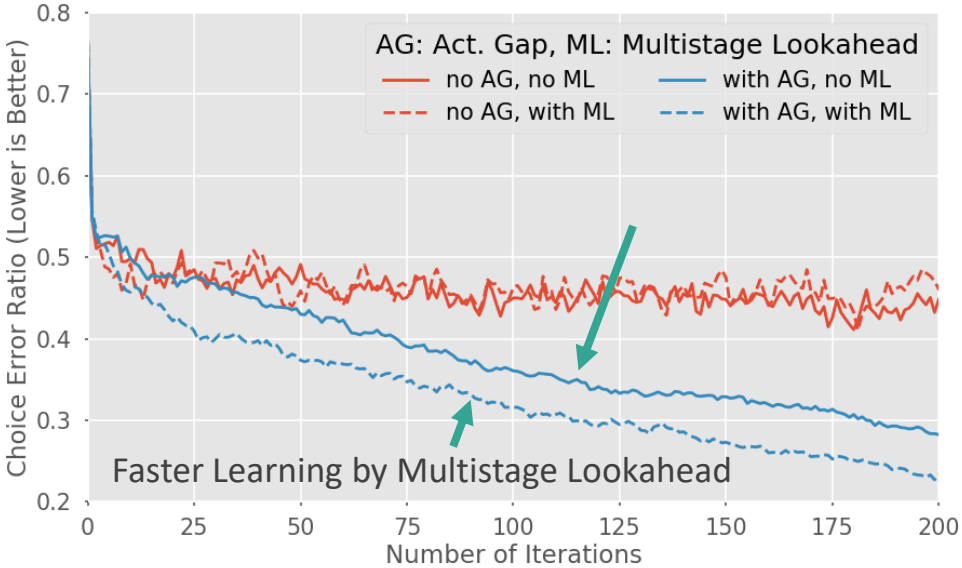
A novel policy evaluation dynamic programming algorithm which is

Efficient (Off-Policy & Multistage Lookahead) and **Robust** to approximation error.

Idea



Result



【学生優秀プレゼンテーション賞対象】

Robust and Efficient Off-Policy Policy Evaluation via Enhanced Action-Gap

Tadashi Kozuno (OIST), Eiji Uchibe (ATR & OIST), Kenji Doya (OIST) ID: D1-37

Problem

We rephrase policy evaluation problem as:

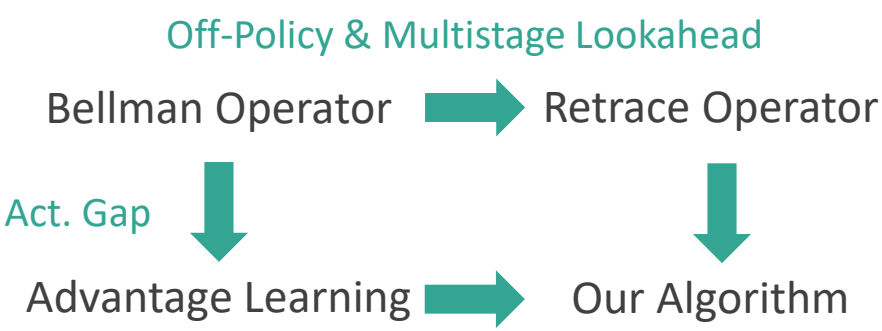
Given a policy π, order actions according to $Q^{\pi}(s,a) \coloneqq \mathbb{E}^{\pi}[\sum_{t\geq 0}\gamma^t R_t \mid s,a]$.

Contributions

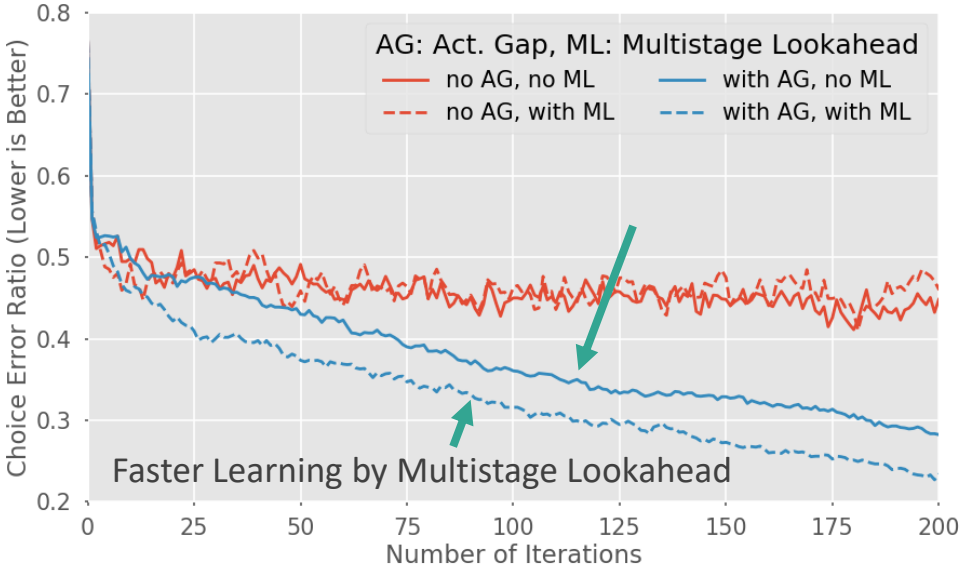
A novel policy evaluation dynamic programming algorithm which is

Efficient (Off-Policy & Multistage Lookahead) and **Robust** to approximation error.

Idea



Result

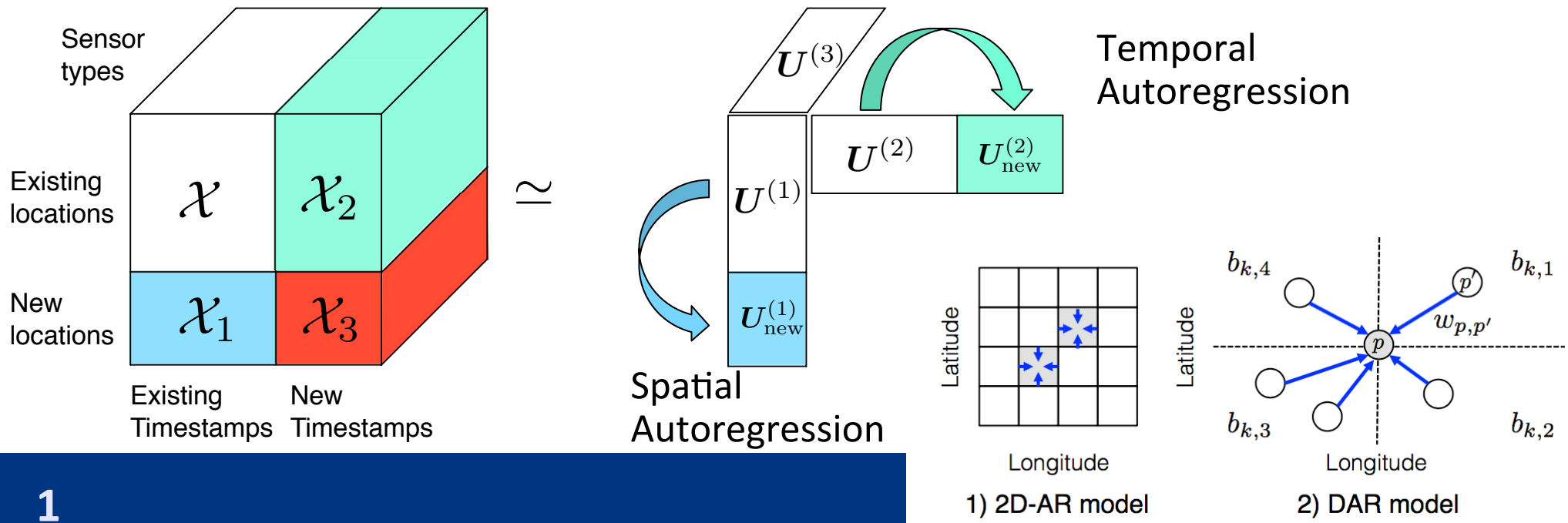


【学生優秀プレゼンテーション賞対象】

D1-38) 自己回帰テンソル分解による時空間予測

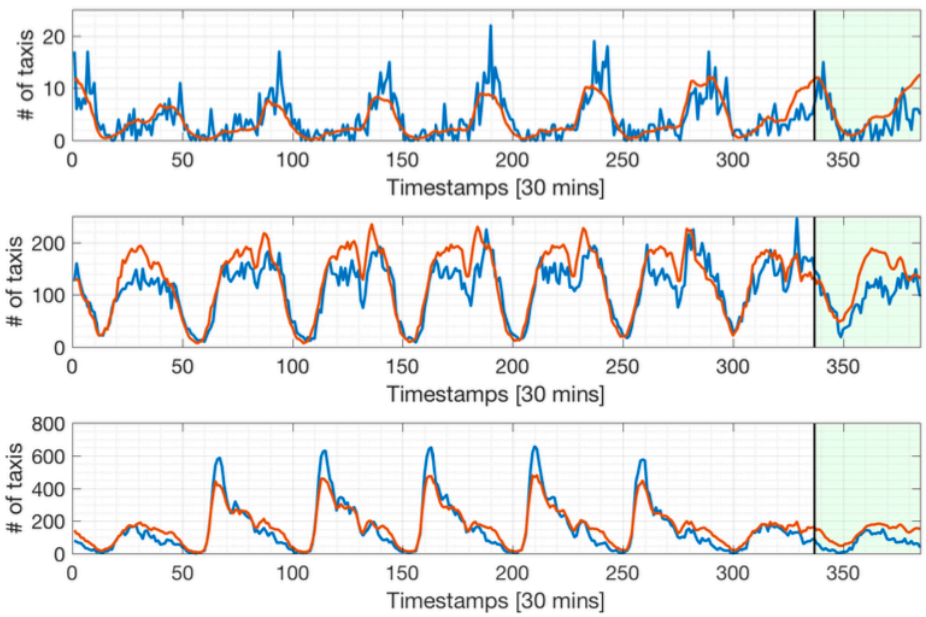
竹内 孝(NTT, 京都大学), 鹿島 久嗣(京都大学, 理研AIP), 上田 修功(NTT, 理研AIP)

- Utilize spatial and temporal graph structures for forecasting future observations on unknown locations.
- Simultaneously decompose a tensor and learn autoregressive models.

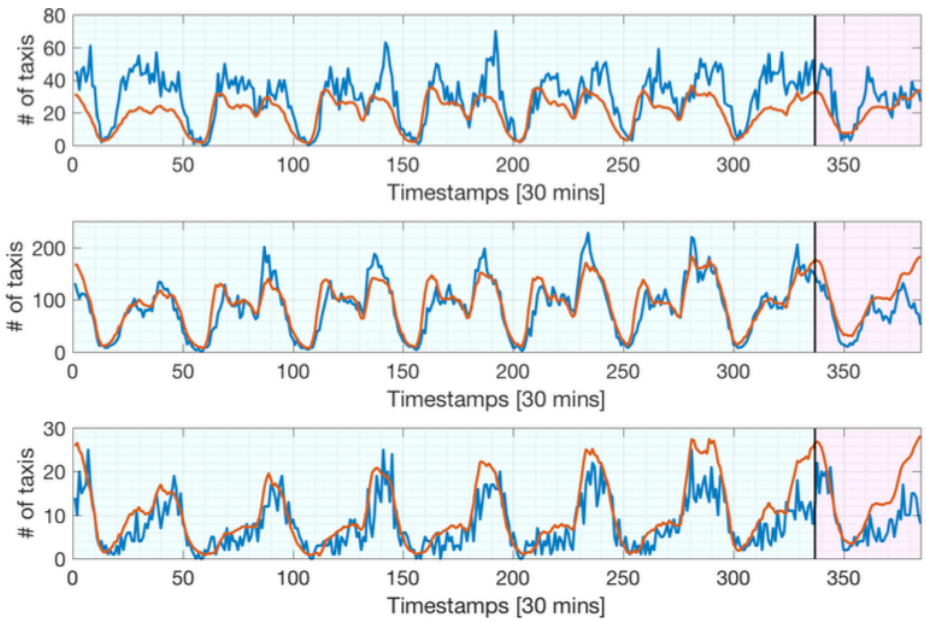


Our proposed method successfully extracted dynamics in Taxi demand forecasting problem

- Green: predictions on known location
  Blue: interpolations on unknown location
  Red: predictions on unknown location (most difficult!)



(a) Predictions of known locations



(b) Predictions and interpolations of unknown locations

On Estimation of Conditional Mode Using Multiple Quantile Regressions

分位点回帰を用いた高次元の条件付き密度・最頻値の推定

Hirofumi Ohta [1] Satoshi Hara [2]

[1]The University of Tokyo, Graduate School of Economics, Department of Statistics, M2.

[2]Osaka University, The Institute of Scientific and Industrial Research.

2017/11/09, IBISML, **ID:D1-39**.

Keywords: conditional mode estimation, quantile regression, dimensionality reduction.

Problem

- Output Y is a scalar and input X is a p-dimensional vector.
- Estimation of conditional density and conditional mode of Y given X.
- For dimensionality reduction in the conditional mode estimation, how to model the distribution of $Y|X$?

Method

- First, we estimate the conditional density by solving linear quantile regressions multiple times.
- We then estimate the conditional mode by finding the maximum of the estimated conditional density.

Contribution

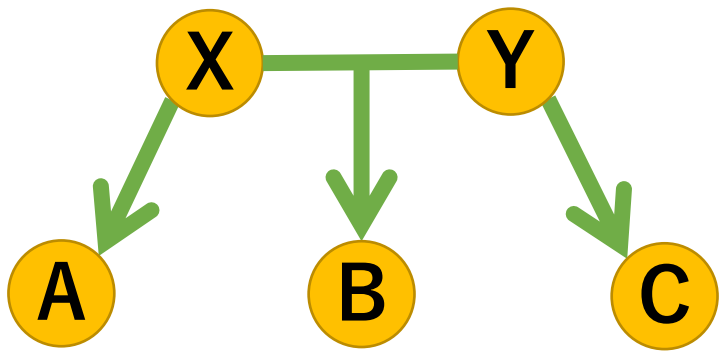
- The proposed method has two advantages:
    - It is computationally stable because it has no initial parameter dependencies.
    - It is statistically efficient with a fast convergence rate.
- Synthetic and real-world data experiments demonstrate the better performance of the proposed method compared to other existing methods.

Probability Tensor Networks

D1-40

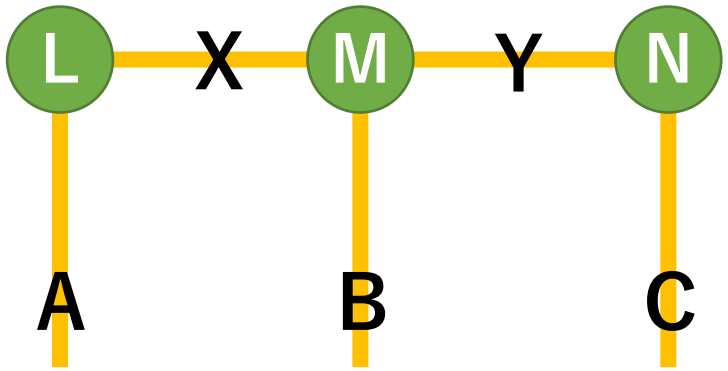
Masakazu ISHIHATA (Hokkaido University), Kohei HAYASHI (AIST/AIP)

Bayesian Networks



$\subset$

Probability Tensor Networks



Conditional Probability Tables

$$p(A, B, C, X, Y) = p(X)p(A \mid X)\; p(Y)p(B \mid X, Y)\; p(C \mid Y)$$

Random Variables

$p(X)p(A \mid X)$	$p(Y)p(B \mid X, Y)$	$p(C \mid Y)$
L	M	N
$L_{a,x}=p(x)p(a\mid x)$	$M_{b,x,y} = p(y)p(b\mid x,y)$	$N_{c,y} = p(c\mid y)$

Probability Tensors

What is PTN? | How different from BN? | When is PTN useful?

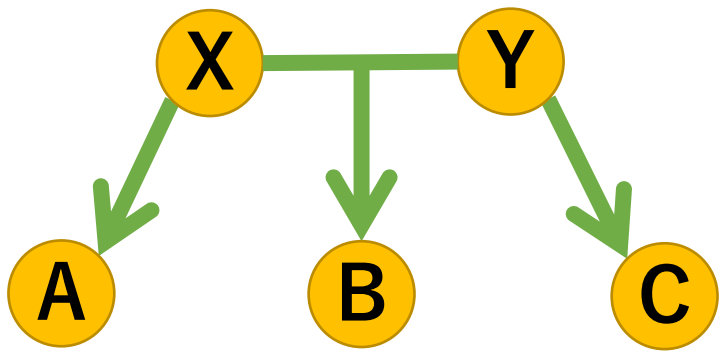
Check our poster!

Probability Tensor Networks

D1-40

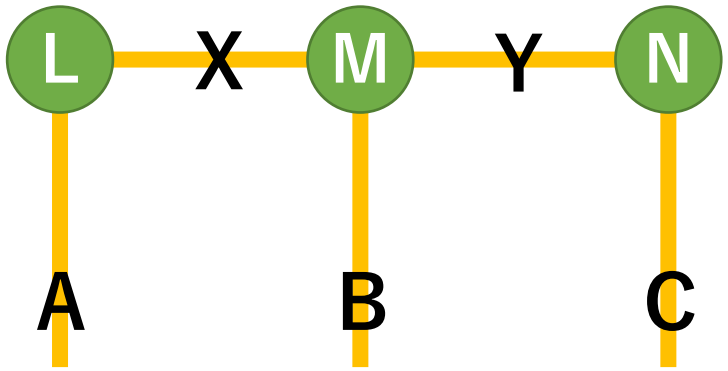
Masakazu ISHIHATA (Hokkaido University), Kohei HAYASHI (AIST/AIP)

Bayesian Networks



$\subset$

Probability Tensor Networks



Conditional Probability Tables

$$p(A, B, C, X, Y) = p(X)p(A \mid X)\; p(Y)p(B \mid X, Y)\; p(C \mid Y)$$

Random Variables

$p(X)p(A \mid X)$	$p(Y)p(B \mid X, Y)$	$p(C \mid Y)$
L	M	N
$L_{a,x}=p(x)p(a\mid x)$	$M_{b,x,y} = p(y)p(b\mid x,y)$	$N_{c,y} = p(c\mid y)$

Probability Tensors

What is PTN? | How different from BN? | When is PTN useful?

Check our poster!

D1-41 学生優秀プレゼンテーション賞対象

IILasso: 相関情報を罰則項に導入したスパースモデリング

髙田正彬(総合研究大学院大学/株式会社東芝), 鈴木大慈(東京大学), 藤澤洋徳(統計数理研究所)

相関情報を利用するスパースモデリングIILassoを提案

Independently Interpretable Lasso

$$\min_{\beta} \frac{1}{2n}\|y - X\beta\|_2^2 + \lambda\left(\|\beta\|_1 + \frac{\alpha}{2}|\beta|^\top R|\beta|\right)$$

$X \in \mathbb{R}^{n\times p}$ ：説明変数, $y \in \mathbb{R}^n$ ：目的変数, $\beta \in \mathbb{R}^p$ ：回帰係数,

$R \in \mathbb{R}^{p\times p}$ ：相関の強さを表す行列, $\lambda > 0, \alpha > 0$ ：正則化パラメータ

Lassoの問題点

互いに高相関の変数を選択してしまう

- X モデルの解釈性が悪化する
- X 汎化性能が低下する

提案手法のメリット

高相関の変数を選択することを抑制

- ○ **モデルの解釈性が向上する**
- ○ **過学習しにくくなる**

理論解析	**符号一致条件**と**収束レート**を導出、有意性を示した
数値実験	**実データ10個で評価**、既存手法より優れていることを示した

D1-41 学生優秀プレゼンテーション賞対象

IILasso: 相関情報を罰則項に導入したスパースモデリング

髙田正彬(総合研究大学院大学/株式会社東芝), 鈴木大慈(東京大学), 藤澤洋徳(統計数理研究所)

相関情報を利用するスパースモデリングIILassoを提案

Independently Interpretable Lasso

$$\min_{\beta} \frac{1}{2n}\|y - X\beta\|_2^2 + \lambda\left(\|\beta\|_1 + \frac{\alpha}{2}|\beta|^\top R|\beta|\right)$$

$X \in \mathbb{R}^{n\times p}$ ：説明変数, $y \in \mathbb{R}^n$ ：目的変数, $\beta \in \mathbb{R}^p$ ：回帰係数,

$R \in \mathbb{R}^{p\times p}$ ：相関の強さを表す行列, $\lambda > 0, \alpha > 0$ ：正則化パラメータ

Lassoの問題点

互いに高相関の変数を選択してしまう

- X モデルの解釈性が悪化する
- X 汎化性能が低下する

提案手法のメリット

高相関の変数を選択することを抑制

- ○ **モデルの解釈性が向上する**
- ○ **過学習しにくくなる**

理論解析	**符号一致条件**と**収束レート**を導出、有意性を示した
数値実験	**実データ10個で評価**、既存手法より優れていることを示した

D1-42: 深層モデルを用いたDisentangleな表現学習

NTT 山田真徳, 岩田具治

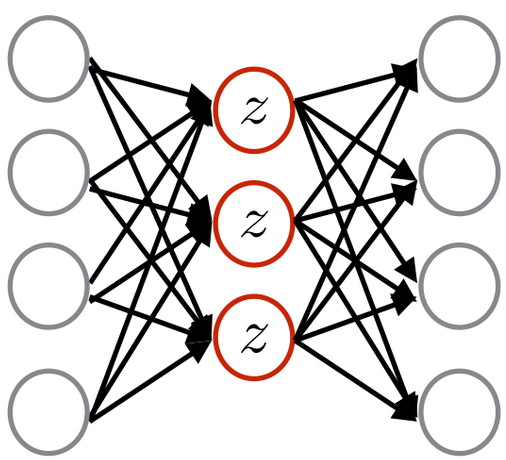
ポスターの内容

- β-VAEとInfoGANの理論的関係を示す
- 相互情報量の可視化の結果

基礎となるmodel　　disentanglement 化

VAE (arXiv:1312.6114 [stat.ML]) → β-VAE (arXiv:1606.05579 [stat.ML])

GAN (arXiv:1406.2661 [stat.ML]) → InfoGAN (arXiv:1606.03657 [cs.LG])



Disentangle表現とは

1つのzのノードで1つの意味を表現

InfoGAN

$$\min_{G,Q}\max_{D}\Big[\underbrace{E_{x\sim p_{data}(x)}\left[\log D\left(x\right)\right]+E_{x\sim G(z,c)}\left[\log\left(1-D\left(x\right)\right)\right]}_{V(G,D)}-\lambda\Big(\overbrace{\underbrace{E_{x\sim G(z,c),c\sim p(c)}\left[\log Q\left(c|x\right)\right]+H\left(c\right)}_{\text{正則化項}}}^{L_I}+E_{x\sim G(z,c)}\left[D_{\mathrm{KL}}\left(P\left(\cdot|x\right)||Q\left(\cdot|x\right)\right)\right]\Big)\Big]$$

$I\left(c;G\left(z,c\right)\right)$

変分下限

β-VAE

$$\max_{\phi,\theta}\log p\left(x\right)=\max_{\phi,\theta}\Big[\underbrace{E_{q_\phi(z|x)}\log p_\theta\left(x|z\right)}_{\text{再構成誤差}}-\underbrace{\beta D_{\mathrm{KL}}\left(q_\phi\left(z|x\right)||p_\theta\left(z\right)\right)}_{\text{正則化項}}+D_{\mathrm{KL}}\left(q_\phi\left(z|x\right)||p_\theta\left(z|x\right)\right)\Big]$$

変分下限

ほぼ同じ制約項(青)

InfoGAN

$$\min_{G,Q}\max_{D}\Big[\underbrace{E_{x\sim p_{data}(x)}\left[\log D\left(x\right)\right]+E_{x\sim G(z,c)}\left[\log\left(1-D\left(x\right)\right)\right]}_{V(G,D)}+\lambda E_{x\sim G(z,c)}\left[D_{\mathrm{KL}}\left(p\left(c\right)||Q\left(c|x\right)\right)\right]\Big]$$

β-VAE

$$\min_{\phi,\theta}\Big[\underbrace{-E_{q_\phi(z|x)}\log p_\theta\left(x|z\right)}_{\text{再構成誤差}}+\beta D_{\mathrm{KL}}\left(q_\phi\left(z|x\right)||p_\theta\left(z\right)\right)\Big]$$

β-VAEとInfoGANは以下の違い→ $D_{\mathrm{KL}}\left(q_\phi\left(z|x\right)||p\left(z\right)\right)\Longleftrightarrow D_{\mathrm{KL}}\left(p\left(z\right)||q_\phi\left(z|x\right)\right)$

NTT

Copyright©2017 NTT corp. All Rights Reserved.

D1-42: 深層モデルを用いたDisentangleな表現学習

NTT 山田真徳, 岩田具治

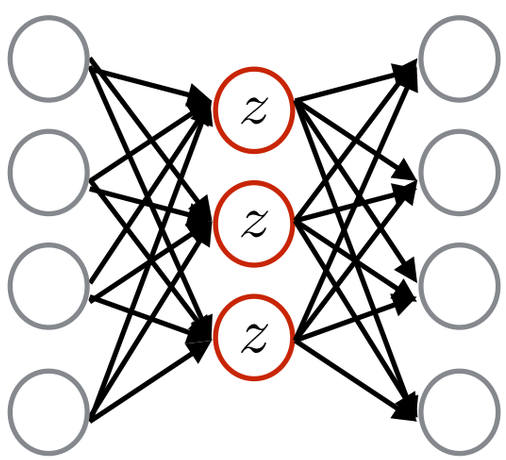
ポスターの内容

- β-VAEとInfoGANの理論的関係を示す
- 相互情報量の可視化の結果

基礎となるmodel　　disentanglement 化

VAE (arXiv:1312.6114 [stat.ML]) → β-VAE (arXiv:1606.05579 [stat.ML])

GAN (arXiv:1406.2661 [stat.ML]) → InfoGAN (arXiv:1606.03657 [cs.LG])



Disentangle表現とは

1つのzのノードで1つの意味を表現

InfoGAN

$$\min_{G,Q}\max_{D}\Big[\underbrace{E_{x\sim p_{data}(x)}\left[\log D\left(x\right)\right]+E_{x\sim G(z,c)}\left[\log\left(1-D\left(x\right)\right)\right]}_{V(G,D)}-\lambda\Big(\overbrace{\underbrace{E_{x\sim G(z,c),c\sim p(c)}\left[\log Q\left(c|x\right)\right]+H\left(c\right)}_{\text{正則化項}}}^{L_I}+E_{x\sim G(z,c)}\left[D_{\mathrm{KL}}\left(P\left(\cdot|x\right)||Q\left(\cdot|x\right)\right)\right]\Big)\Big]$$

$I\left(c;G\left(z,c\right)\right)$

変分下限

β-VAE

$$\max_{\phi,\theta}\log p\left(x\right)=\max_{\phi,\theta}\Big[\underbrace{E_{q_\phi(z|x)}\log p_\theta\left(x|z\right)}_{\text{再構成誤差}}-\underbrace{\beta D_{\mathrm{KL}}\left(q_\phi\left(z|x\right)||p_\theta\left(z\right)\right)}_{\text{正則化項}}+D_{\mathrm{KL}}\left(q_\phi\left(z|x\right)||p_\theta\left(z|x\right)\right)\Big]$$

変分下限

ほぼ同じ制約項(青)

InfoGAN

$$\min_{G,Q}\max_{D}\Big[\underbrace{E_{x\sim p_{data}(x)}\left[\log D\left(x\right)\right]+E_{x\sim G(z,c)}\left[\log\left(1-D\left(x\right)\right)\right]}_{V(G,D)}+\lambda E_{x\sim G(z,c)}\left[D_{\mathrm{KL}}\left(p\left(c\right)||Q\left(c|x\right)\right)\right]\Big]$$

β-VAE

$$\min_{\phi,\theta}\Big[\underbrace{-E_{q_\phi(z|x)}\log p_\theta\left(x|z\right)}_{\text{再構成誤差}}+\beta D_{\mathrm{KL}}\left(q_\phi\left(z|x\right)||p_\theta\left(z\right)\right)\Big]$$

β-VAEとInfoGANは以下の違い→ $D_{\mathrm{KL}}\left(q_\phi\left(z|x\right)||p\left(z\right)\right)\Longleftrightarrow D_{\mathrm{KL}}\left(p\left(z\right)||q_\phi\left(z|x\right)\right)$

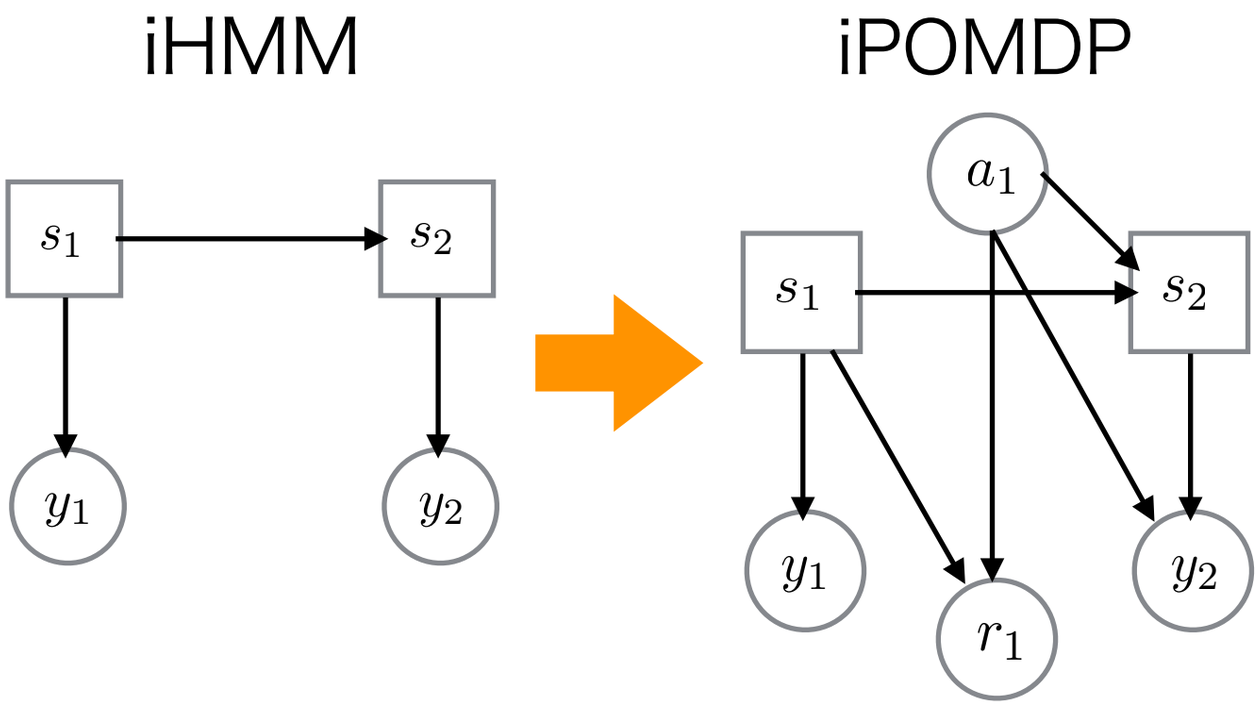
NTT

Copyright©2017 NTT corp. All Rights Reserved.

D1-43：部分観測マルコフ決定過程の環境推定

奥戸道子 (東京大学・情報理工)，森村哲郎 (IBM東京基礎研究所)

POMDP推定に無限隠れマルコフモデル(iHMM)に対するMCMC法を適用する



部分観測マルコフ決定過程 (POMDP)

- 表現力の高い時系列モデル
- 強化学習など多くの意思決定過程のモデリングに用いられる
- POMDPの推定は困難

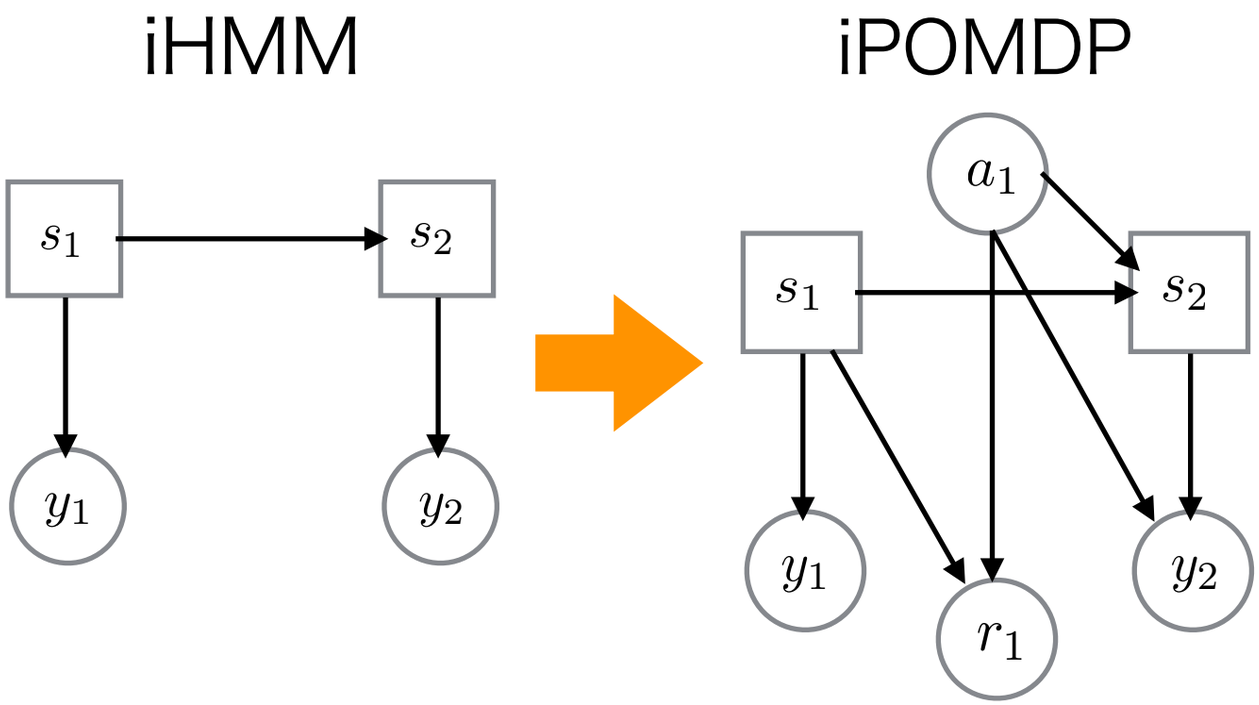
本研究の内容

- POMDPベンチマーク問題で性能を確認.
- 内部状態を作る必要のある問題に対してMCMCの収束が悪いことを実験で確認し，Parallel Temperingの適用を提案.

D1-43：部分観測マルコフ決定過程の環境推定

奥戸道子 (東京大学・情報理工)，森村哲郎 (IBM東京基礎研究所)

POMDP推定に無限隠れマルコフモデル(iHMM)に対するMCMC法を適用する



部分観測マルコフ決定過程 (POMDP)

- 表現力の高い時系列モデル
- 強化学習など多くの意思決定過程のモデリングに用いられる
- POMDPの推定は困難

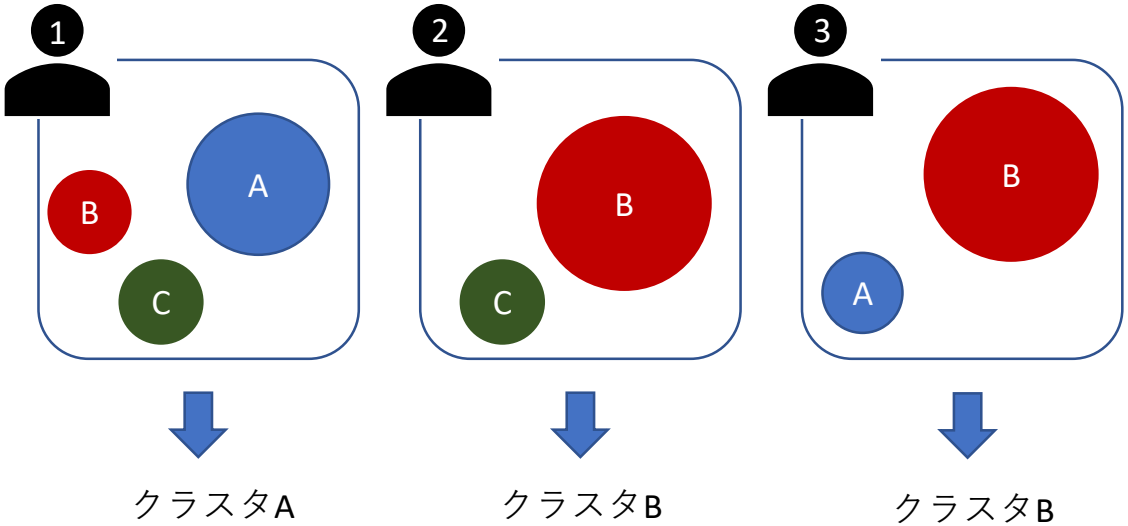
本研究の内容

- POMDPベンチマーク問題で性能を確認.
- 内部状態を作る必要のある問題に対してMCMCの収束が悪いことを実験で確認し，Parallel Temperingの適用を提案.

D1-44: Latent Dirichlet Allocationを用いた腸内細菌叢のコミュニティ解析

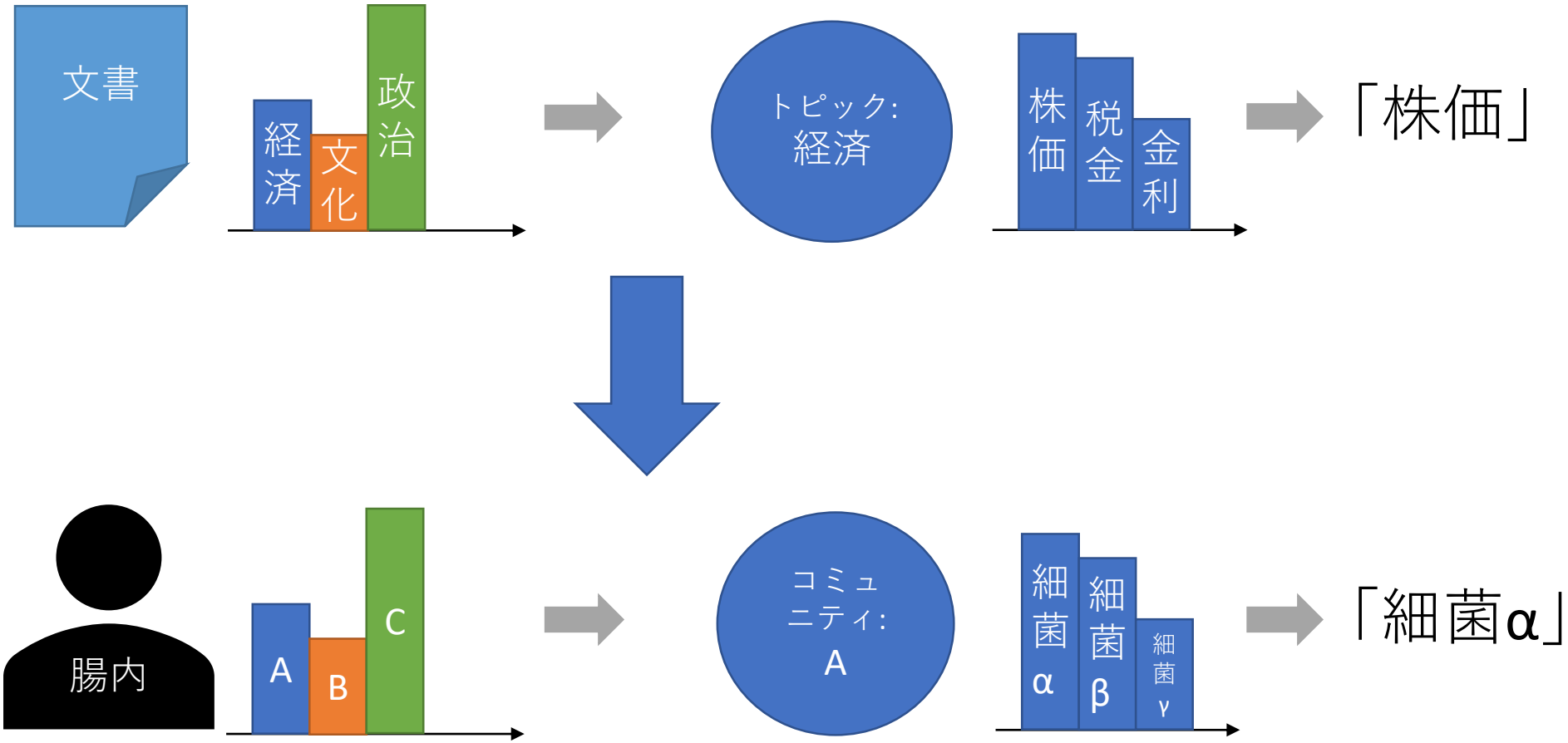
細田至温[1,2]、西嶋傑[2,3,5]、福永津嵩[1,4]、服部正平[5,3]、浜田道昭[1,2,6]
[1]早稲田大学大学院 先進理工学研究科 電気・情報生命専攻
[2]産総研・早大 生体システムビッグデータ解析オープンイノベーションラボラトリ（CBBD-OIL）
[3]東京大学大学院 新領域創成科学研究科
[4]東京大学大学院情報理工学系研究科コンピュータ科学専攻
[5]早稲田大学 理工学術院 先進理工学研究科
[6]日本医科大学 医学研究科

- 腸内細菌はコミュニティ単位で共起している[Boon E et. al., 2014]
  - コミュニティとは細菌のグループ
- 今までは腸内細菌の解析としてクラスタリングが行われてきたが下図のようにコミュニティの推定には不適切



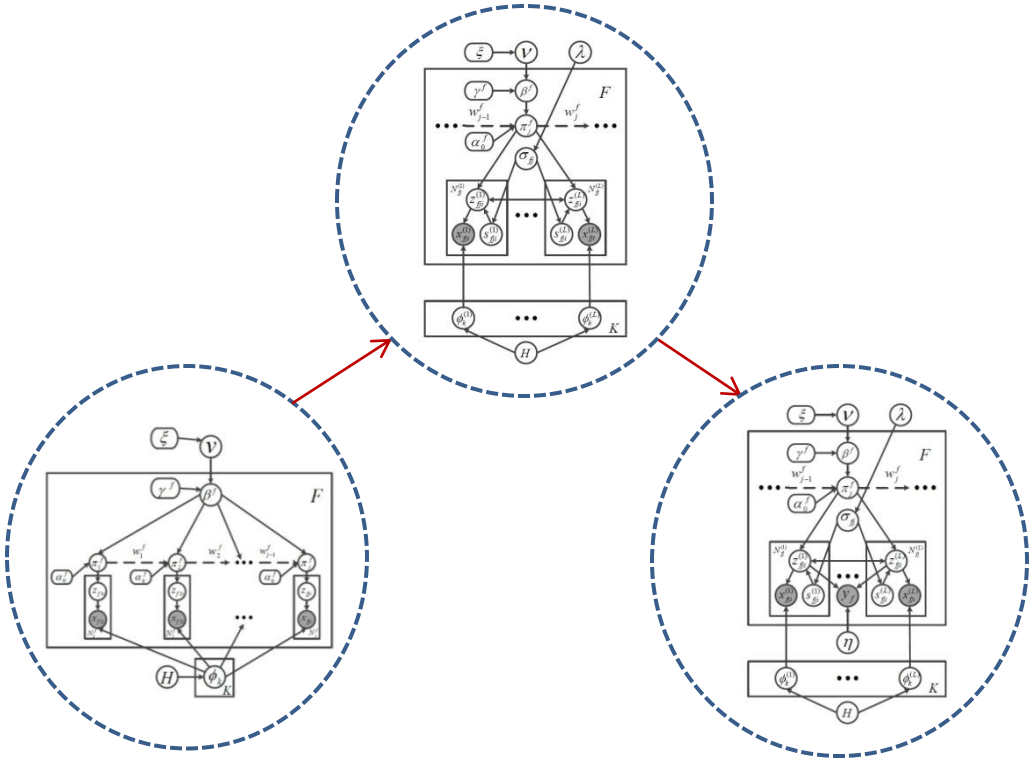
三人の被験者の腸内を表した図。円はコミュニティを指す。

LDAにおける文書-トピック-単語の関係を腸内細菌における腸内-コミュニティ-細菌の関係に当てはめて推定



Supervised Nonparametric Multimodal Topic Modeling Methods for Multi-class Video Classification

Jianfei Xue* and Koji Eguchi
Graduate School of System Informatics, Kobe University



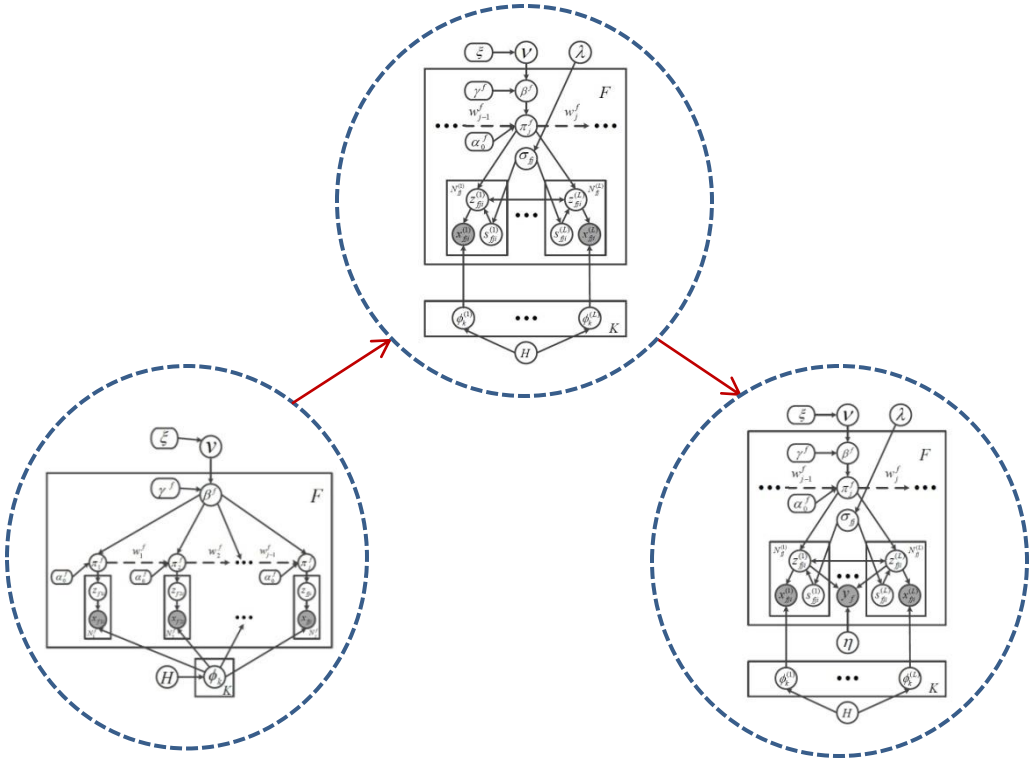
SPOTLIGHT

- Propose a supervised nonparametric topic model to classify multi-class videos.
- Incorporate a time-dependency mechanism into the supervised nonparametric topic model for modeling neighboring video frames.
- Incorporate a more flexible symmetric correspondence method to handle video multimodal data enriched with image data and speech data within the supervised nonparametric topic model.

*「学生優秀プレゼンテーション賞対象」

Supervised Nonparametric Multimodal Topic Modeling Methods for Multi-class Video Classification

Jianfei Xue* and Koji Eguchi
Graduate School of System Informatics, Kobe University



SPOTLIGHT

- Propose a supervised nonparametric topic model to classify multi-class videos.
- Incorporate a time-dependency mechanism into the supervised nonparametric topic model for modeling neighboring video frames.
- Incorporate a more flexible symmetric correspondence method to handle video multimodal data enriched with image data and speech data within the supervised nonparametric topic model.

*「学生優秀プレゼンテーション賞対象」

D1-46 カーネル密度関数の局所変形に基づく線状構造物に対する非剛体イメージアライメント

国立大学法人 九州工業大学

綿島正剛(九工大),久下小百合(九大),石原健(九大),飯野雄一(東大),吉田亮(統数研),徳永旭将(九工大)

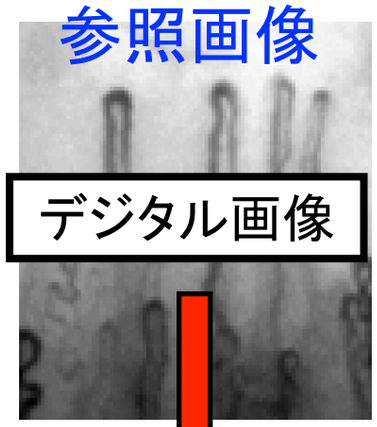
■ バイオイメージアライメント技術

・異なる条件下で計測された2種類以上の画像を共通の座標系に変換する技術
・血管や網膜、神経突起など、線状構造物のアライメントは難易度が高い問題

■ 提案手法

・カーネル密度推定 ＋ EMアルゴリズムによるロバストな非剛体補正を実現

参照画像



▶カーネル密度変換(連続関数に変換)

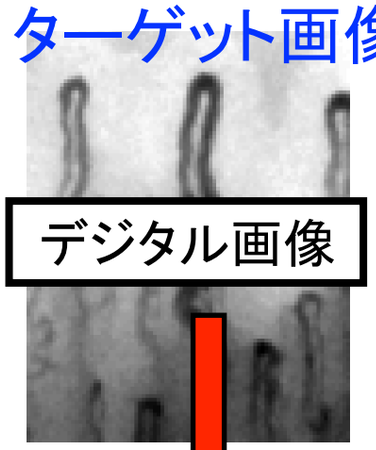
$$\begin{cases} \text{参照画像}: & p_s(x)=\sum_{j=1}^{n} w_j^s \delta(x-x_j) \\ \text{ターゲット画像}: & p_f(x)=\sum_{i=1}^{n} w_i^f k\left(x-\varphi_i^{\mu^f}\right) \end{cases}$$

▶最適化:KLD最小化

$$\hat{\varphi}^{\mu^f} = \arg\min_{\hat{\varphi}^{\mu^f}} \mathrm{KLD}\left(p_s \,\|\, p_f(x \mid \hat{\varphi}^{\mu^f})\right)$$

自由変形を含む最適な対応関係を推定

ターゲット画像



D1-46 カーネル密度関数の局所変形に基づく線状構造物に対する非剛体イメージアライメント

国立大学法人 九州工業大学
綿島正剛(九工大),久下小百合(九大),石原健(九大),飯野雄一(東大),吉田亮(統数研),徳永旭将(九工大)

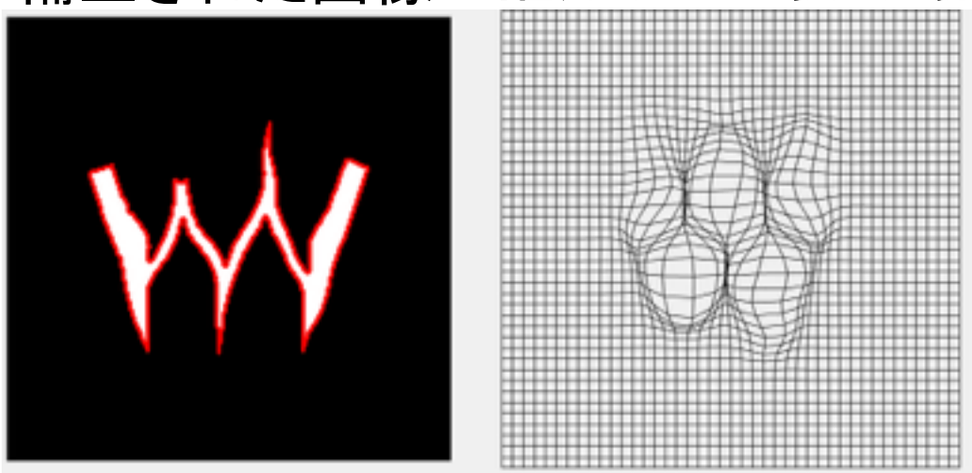
参照画像　ターゲット画像

実験

W字型のターゲット画像を補正して、長方形型の参照画像にどこまで近づけることができるか？

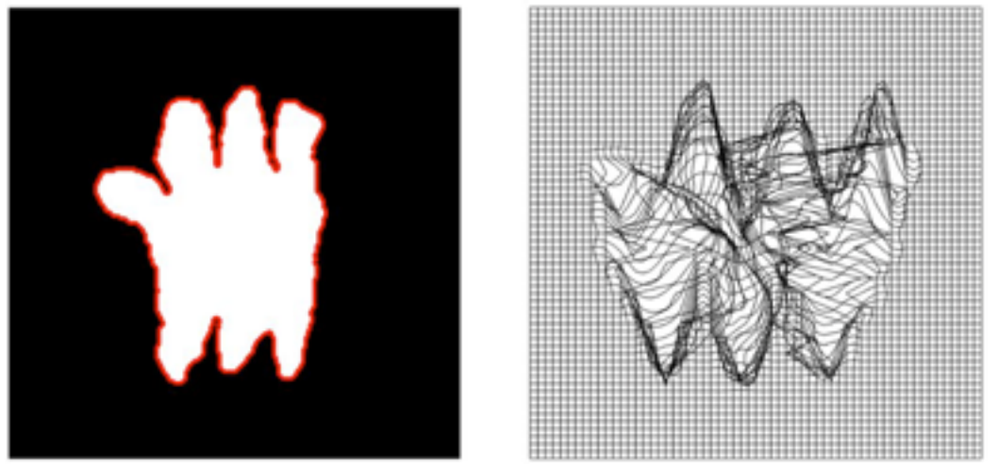
先行研究

補正された画像　推定された変形場



提案手法

補正された画像　推定された変形場



■ 提案手法では、対応する特徴点がなくてもターゲット画像を参照画像に”強引に”近づけることができている(大変形の補正も可能)

交換Monte Carlo法を用いた MR画像事前分布の正則化パラメータ推定の検証

松尾篤樹，原田賢，井上真郷（早稲田大学）

➢ 背景

MR画像再構成におけるグリッドサーチやベイズ最適化による正則化パラメータλの推定は最適化計算が繰り返し必要なため計算量が多い

$$\underset{\boldsymbol{x}}{\operatorname{argmin}}\left\{\frac{1}{2\sigma^2}\sum_c \|\boldsymbol{y}_c - \boldsymbol{F}_u\boldsymbol{S}_c\boldsymbol{x}\|_2^2 + \lambda\|\mathrm{proj}(\boldsymbol{x})\|_{\mathrm{TV}}\right\}$$

➢ 提案

データセットに対する事前分布$P(\boldsymbol{x};\lambda)$の最尤解$\hat{\lambda}$を正則化パラメータとして用いる

- 観測モデルによらない事前分布単独でのパラメータ推定が可能
- 画像データセットから算出でき，観測データは不要

➡ グリッドサーチによるパラメータ推定の結果と比較検証

交換Monte Carlo法を用いた MR画像事前分布の正則化パラメータ推定の検証

松尾篤樹，原田賢，井上真郷（早稲田大学）

➢ 背景

MR画像再構成におけるグリッドサーチやベイズ最適化による正則化パラメータλの推定は最適化計算が繰り返し必要なため計算量が多い

$$\underset{\boldsymbol{x}}{\operatorname{argmin}}\left\{\frac{1}{2\sigma^2}\sum_c \|\boldsymbol{y}_c - \boldsymbol{F}_u\boldsymbol{S}_c\boldsymbol{x}\|_2^2 + \lambda\|\mathrm{proj}(\boldsymbol{x})\|_{\mathrm{TV}}\right\}$$

➢ 提案

データセットに対する事前分布$P(\boldsymbol{x};\lambda)$の最尤解$\hat{\lambda}$を正則化パラメータとして用いる

- 観測モデルによらない事前分布単独でのパラメータ推定が可能
- 画像データセットから算出でき，観測データは不要

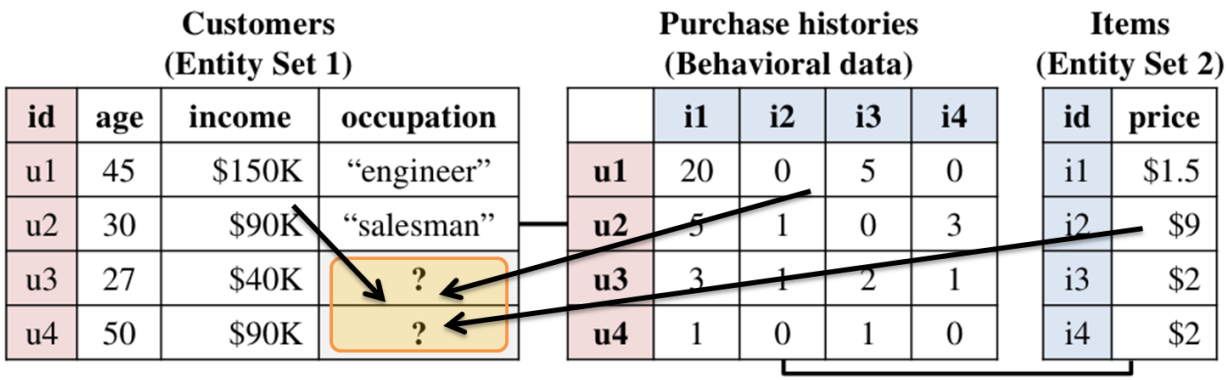
➡ グリッドサーチによるパラメータ推定の結果と比較検証

D1-48: 関係データに向けた混合エキスパートモデル

小山田 昌史 (NEC), 中台 慎二 (NEC)

◆ タスク: 関係データにおけるマスタ表の属性予測

- **(入力)**
  - ✓ **[マスタ表] 顧客の基本的な属性値** (例: 年齢, 性別)
    - – 基本的な属性値は、全顧客について与えられていると仮定
  - ✓ **[ファクト表] 顧客の行動データ** (例: 購買履歴、GPSログ)
    - – 「ユーザ x 商品」のような行列/テンソル
- **(出力)**
  - ✓ **[マスタ表] 顧客の詳細な属性値** (例: 職業)



◆ 課題: いかに「ファクト表（グラフ）」を活用するか

従来のアプローチと、その課題

1.　専門家の知識を使い特徴量エンジニアリング
(例: 職業をを当てるには「平均購買金額」が有効)
→ 専門知識が必要で、時間も人手もかかる

2. ファクト表を特異値分解/テンソル分解で埋め込む
→ 自動生成された特徴は解釈/説明が難しい

id	age	income	avg. purchase	a1	a2	occupation
u1	45	$150K	$1.75	20.6	-0.88	"engineer"
u2	30	$90K	$10.16	4.98	3.15	"salesman"
u3	27	$40K	$8.125	3.44	0.77	?
u4	50	$90K	$1.75	1.20	-0.24	?

Augmented features

本研究のゴール: 特徴量を生成することなく高速・高精度にマスタ表の属性予測を実現

（そのために、関係データ向け混合エキスパートモデルを提案）

D1-49

分解型NML符号長を用いた連続値潜在変数モデルの選択

岡田誠(東大, CREST), 山西健司(東大, CREST)

（学生優秀プレゼンテーション賞対象）

研究背景

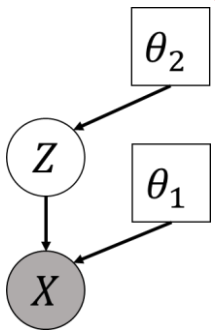
記述長最小原理に基づく潜在変数モデルのモデル選択

２つのアプローチ

- 完全変数化NML符号長
  様々なモデルへの適用事例あり
- 分解型NML符号長 (DNML符号長)
  離散モデルへの適用事例あり

アイデア

(X, Z)の生成をXの生成と
ZからのXの生成の二段階
に分けてNML符号化



利点

- 効率計算可能
- 良いミニマックス最適性

本研究

DNML符号長を連続値潜在変数モデルへ拡張

混合ガウスモデル，確率的主成分分析モデル

実験結果

- 完全変数化NML符号長と同等か，それ以上のモデル推定精度
- 従来手法に比べ，ノイズに対してロバストなモデル選択

D1-49

分解型NML符号長を用いた連続値潜在変数モデルの選択

岡田誠(東大, CREST), 山西健司(東大, CREST)

（学生優秀プレゼンテーション賞対象）

研究背景

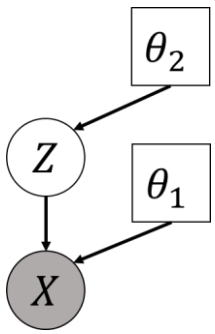
記述長最小原理に基づく潜在変数モデルのモデル選択

２つのアプローチ

- 完全変数化NML符号長
  様々なモデルへの適用事例あり
- 分解型NML符号長 (DNML符号長)
  離散モデルへの適用事例あり

アイデア

(X, Z)の生成をXの生成と
ZからのXの生成の二段階
に分けてNML符号化



利点

- 効率計算可能
- 良いミニマックス最適性

本研究

DNML符号長を連続値潜在変数モデルへ拡張

混合ガウスモデル，確率的主成分分析モデル

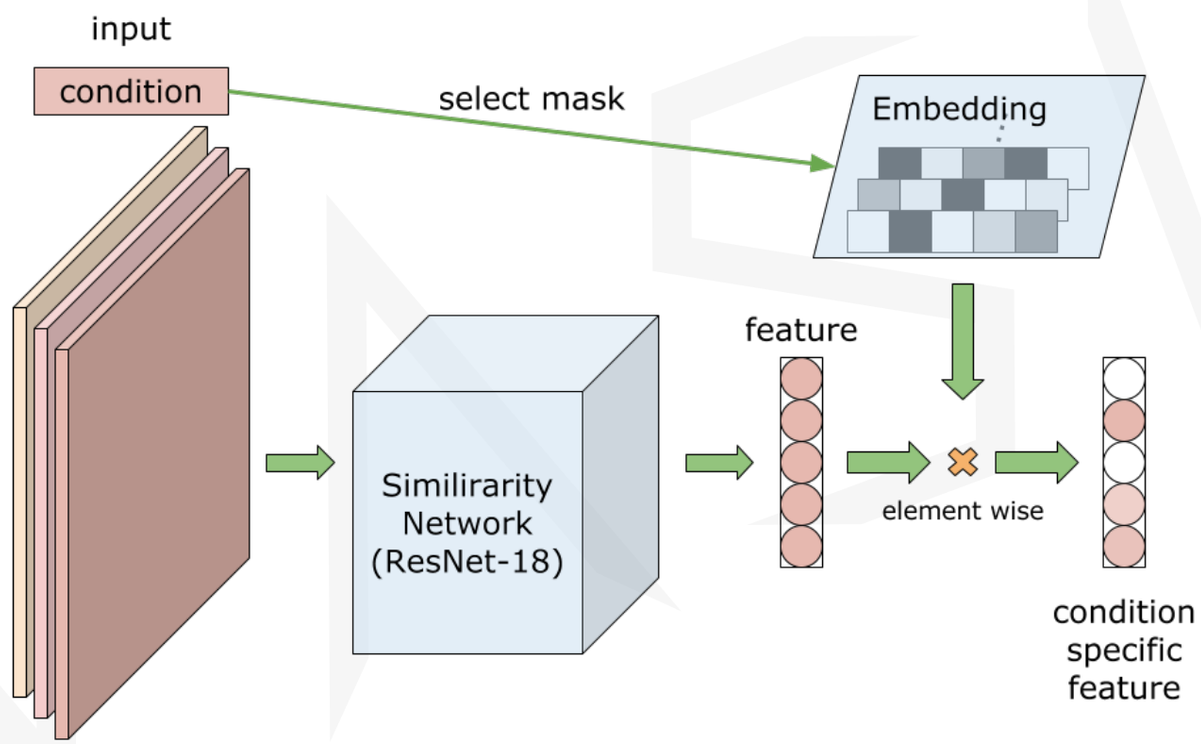
実験結果

- 完全変数化NML符号長と同等か，それ以上のモデル推定精度
- 従来手法に比べ，ノイズに対してロバストなモデル選択

D1-50 学習可能なマスクを用いた柔軟な類似度計算手法

上月正貴 (VASILY, Inc./東京大学), 中村拓磨(VASILY, Inc.) , 後藤亮介(VASILY, Inc.)

一度に複数の基準を考慮した類似度を計算

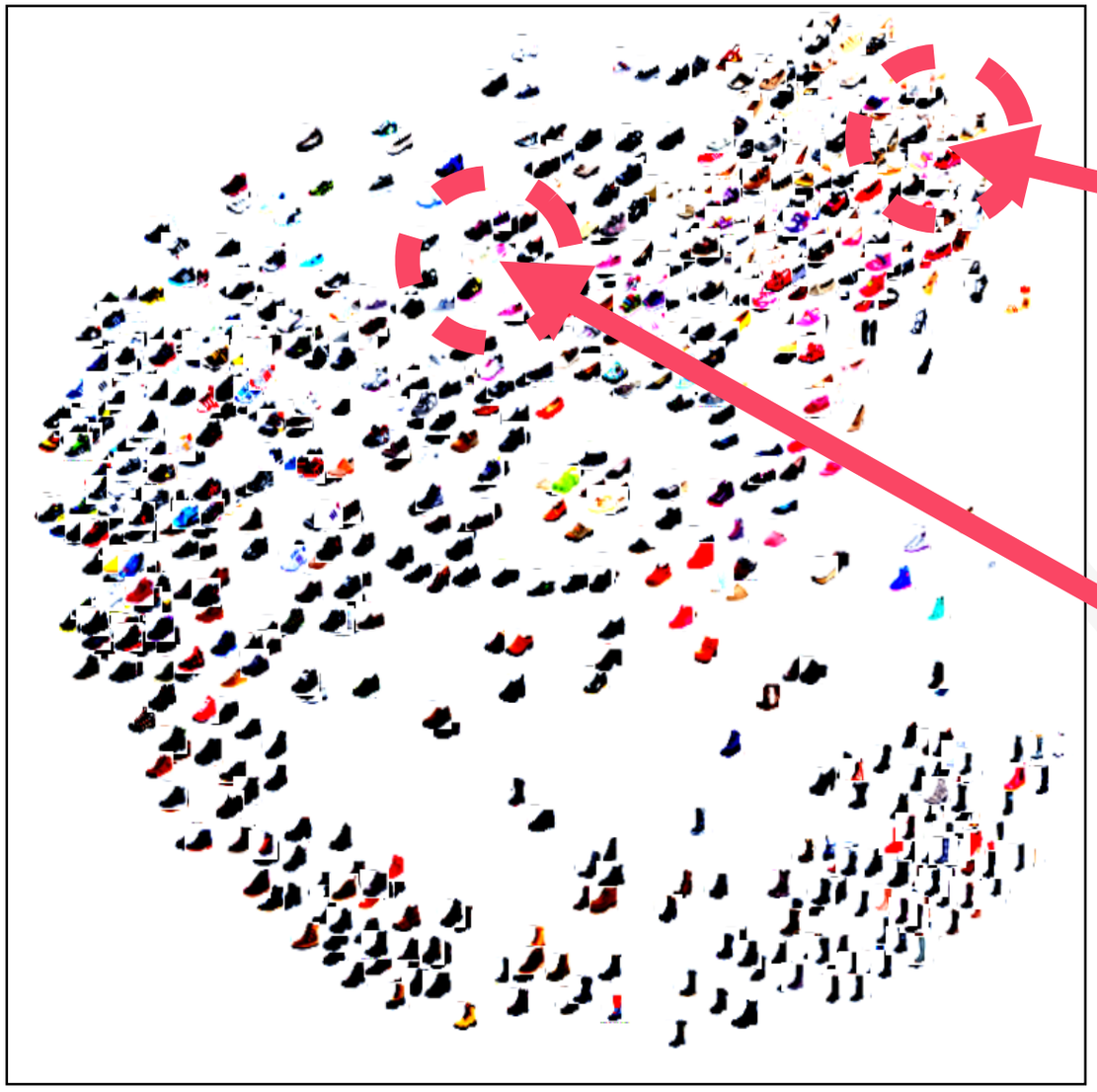


© 2017 VASILY,Inc.

D1-50 学習可能なマスクを用いた柔軟な類似度計算手法

上月正貴 (VASILY, Inc./東京大学), 中村拓磨(VASILY, Inc.) , 後藤亮介(VASILY, Inc.)

靴の種類と性別を考慮して埋込



スニーカー&女の子向け

スポーツシューズ&女性向け

© 2017 VASILY,Inc.

D1-51

学生発表

Training Generative Models with Maximum Mean Discrepancy

崔 正行[1] 前田 新一[2]

[1]東京大学 [2]Preferred Networks

Aim: To estimate a probability by using a valid distance.

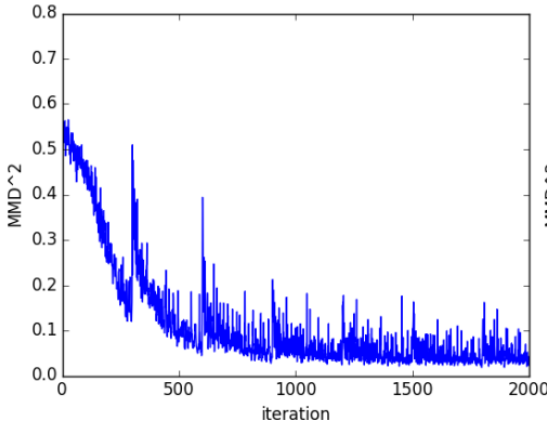
Approach: use Maximum Mean Discrepancy (MMD)

$$\mathrm{MMD}[p(x), q(x)] = \mathbb{E}_{x,x'\sim p(x)}[k(x,x')] + \mathbb{E}_{y,y'\sim q(x)}[k(y,y')]$$
$$-2\mathbb{E}_{x,\sim p(x),y\sim q(x)}[k(x,y)]$$

where $k(\cdot,\cdot)$ is a kernel function

Expectation can be replaced with sampling.

But suffers from high variance

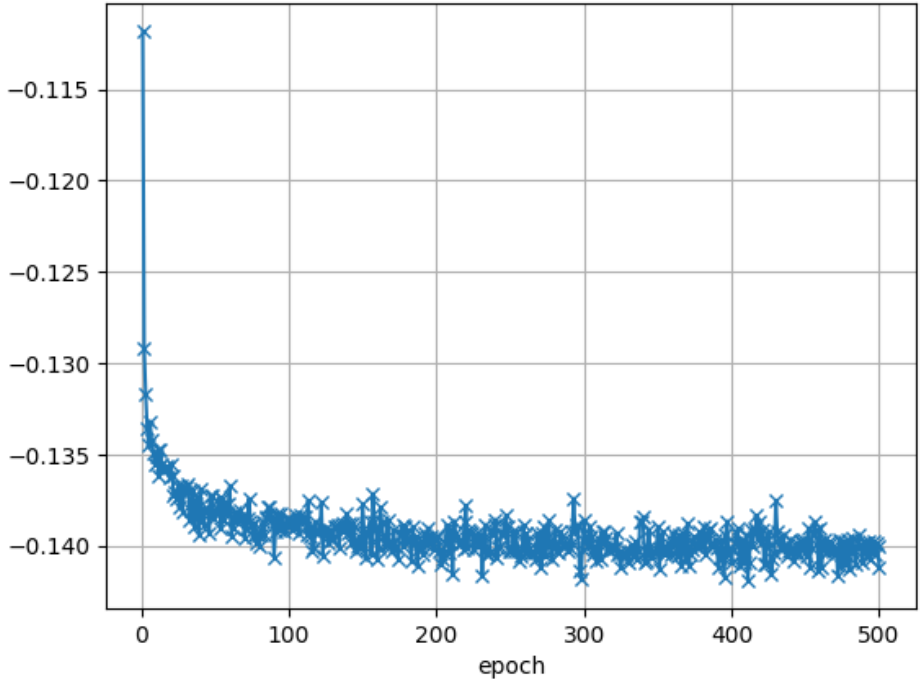


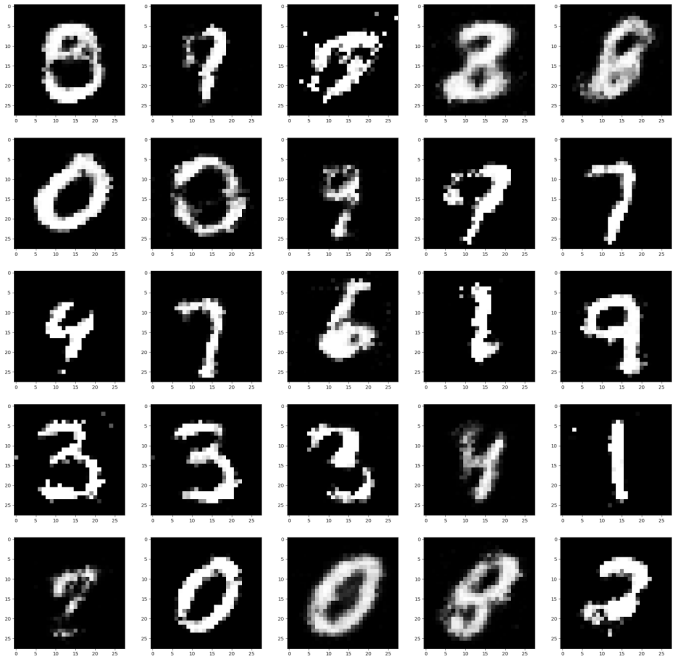
Preferred Networks

Training Generative Models with Maximum Mean Discrepancy

崔 正行[1] 前田 新一[2]
[1]東京大学 [2]Preferred Networks

Remedy: 1. Partially solve the integral analytically





2. Learning a class-conditional generative model (on going)

Preferred Networks

D1-52: Direct Log-Density Gradient Estimation with Gaussian Mixture Models

Qi Zhang, Hiroaki Sasaki, Kazushi Ikeda (NAIST)

Goal: A novel estimator for the gradients of log-densities.

Approach:

- Directly estimate the gradients of log-densities without resorting to density estimation.
- Learn log-density gradient correlation structure by Gaussian mixture models.

Advantage:

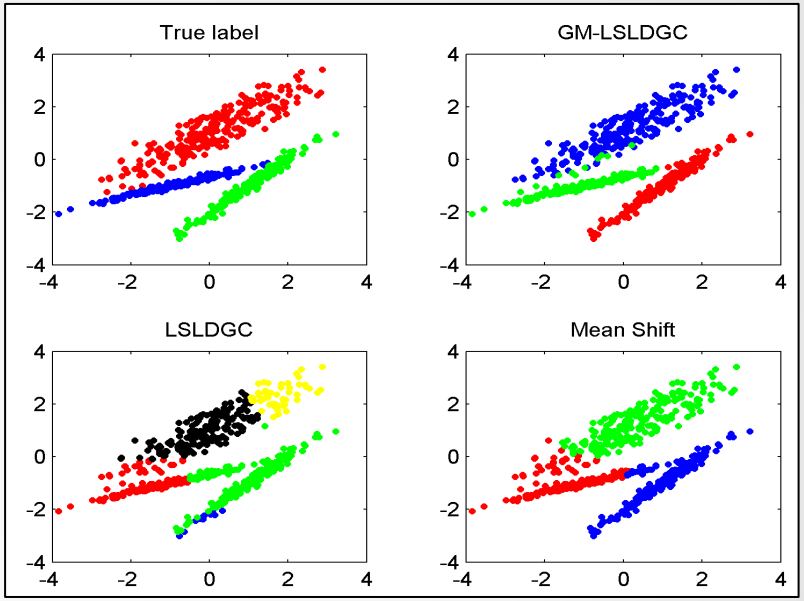
- Auto determines # of clusters.
- Works well when the sample has a high correlation.

Application:

Mode-seeking clustering.

1, Using the proposed method for gradient estimation.

2, A fixed point algorithm for mode-seeking.



学生優秀プレゼンテーション賞対象

Sasaki, Hiroaki, et al. "Mode-Seeking Clustering and Density Ridge Estimation via Direct Estimation of Density-Derivative-Ratios." arXiv preprint arXiv:1707.01711 (2017).APA
Fukunaga, Keinosuke, and Larry Hostetler. "The estimation of the gradient of a density function, with applications in pattern recognition." *IEEE Transactions on information theory* 21.1 (1975): 32-40.

D1-52: Direct Log-Density Gradient Estimation with Gaussian Mixture Models

Qi Zhang, Hiroaki Sasaki, Kazushi Ikeda (NAIST)

Goal: A novel estimator for the gradients of log-densities.

Approach:

- Directly estimate the gradients of log-densities without resorting to density estimation.
- Learn log-density gradient correlation structure by Gaussian mixture models.

Advantage:

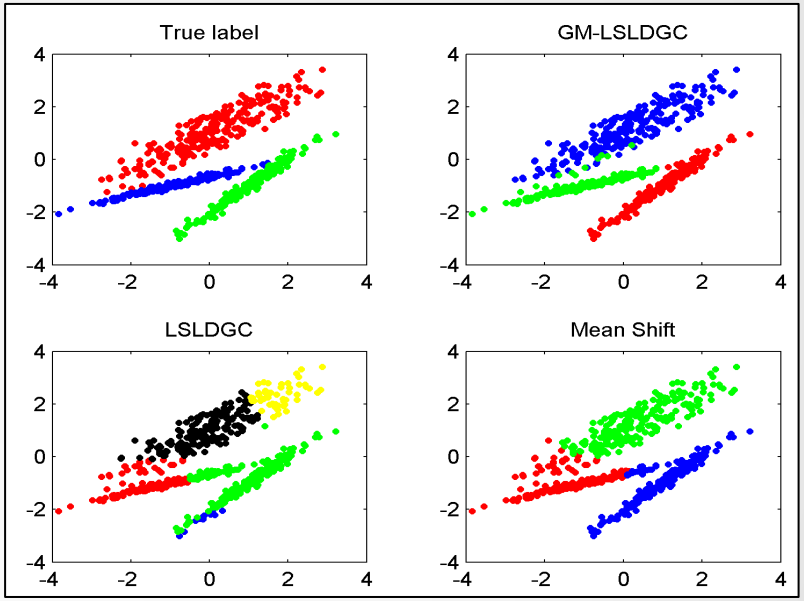
- Auto determines # of clusters.
- Works well when the sample has a high correlation.

Application:

Mode-seeking clustering.

1, Using the proposed method for gradient estimation.

2, A fixed point algorithm for mode-seeking.

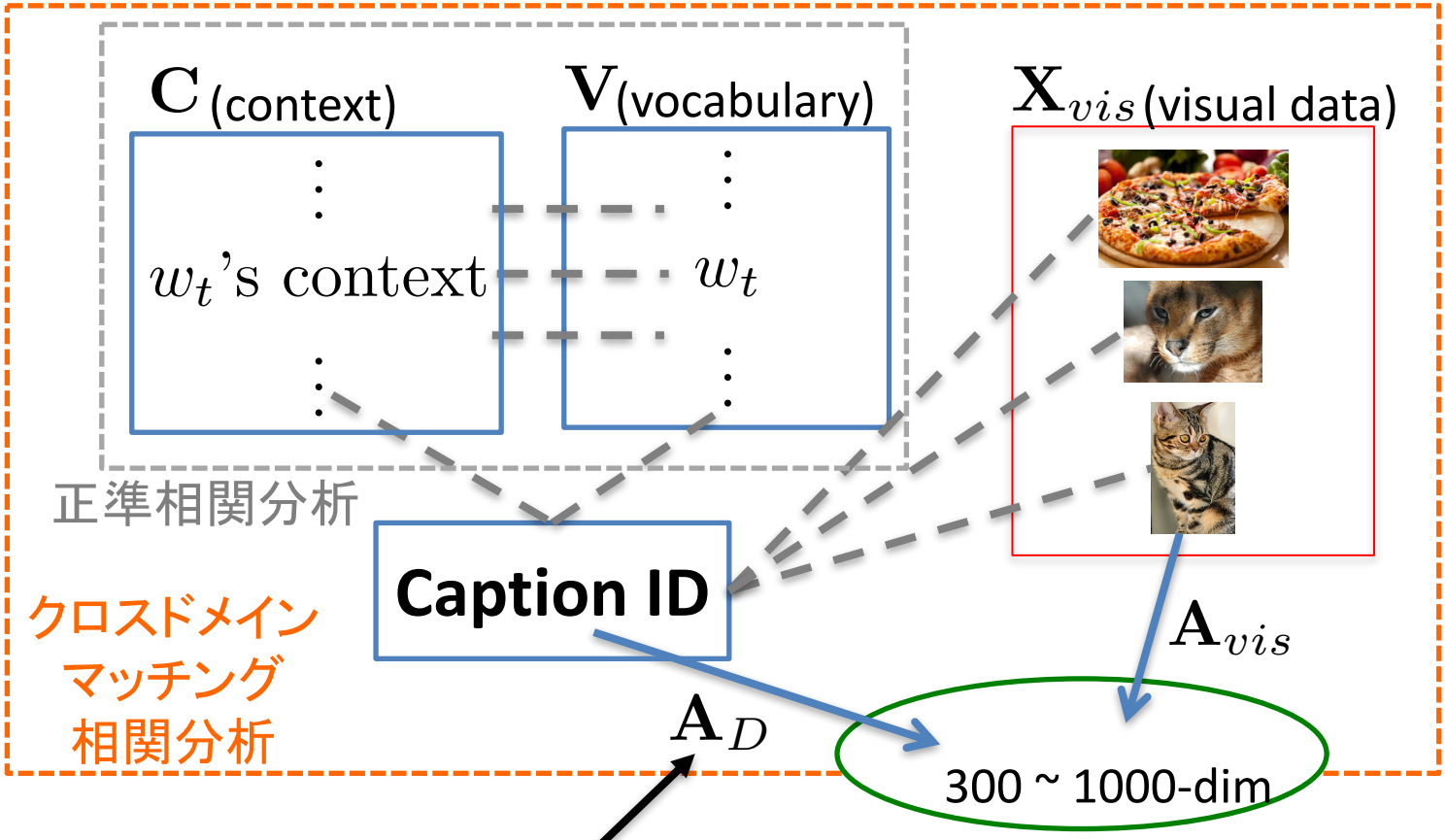


学生優秀プレゼンテーション賞対象

Sasaki, Hiroaki, et al. "Mode-Seeking Clustering and Density Ridge Estimation via Direct Estimation of Density-Derivative-Ratios." arXiv preprint arXiv:1707.01711 (2017).APA
Fukunaga, Keinosuke, and Larry Hostetler. "The estimation of the gradient of a density function, with applications in pattern recognition." *IEEE Transactions on information theory* 21.1 (1975): 32-40.

D1-53：クロスドメインマッチング相関分析を用いた画像とキャプションの同時埋め込み

京都大学大学院情報学研究科 / 理研AIP
福井一輝，下平英寿

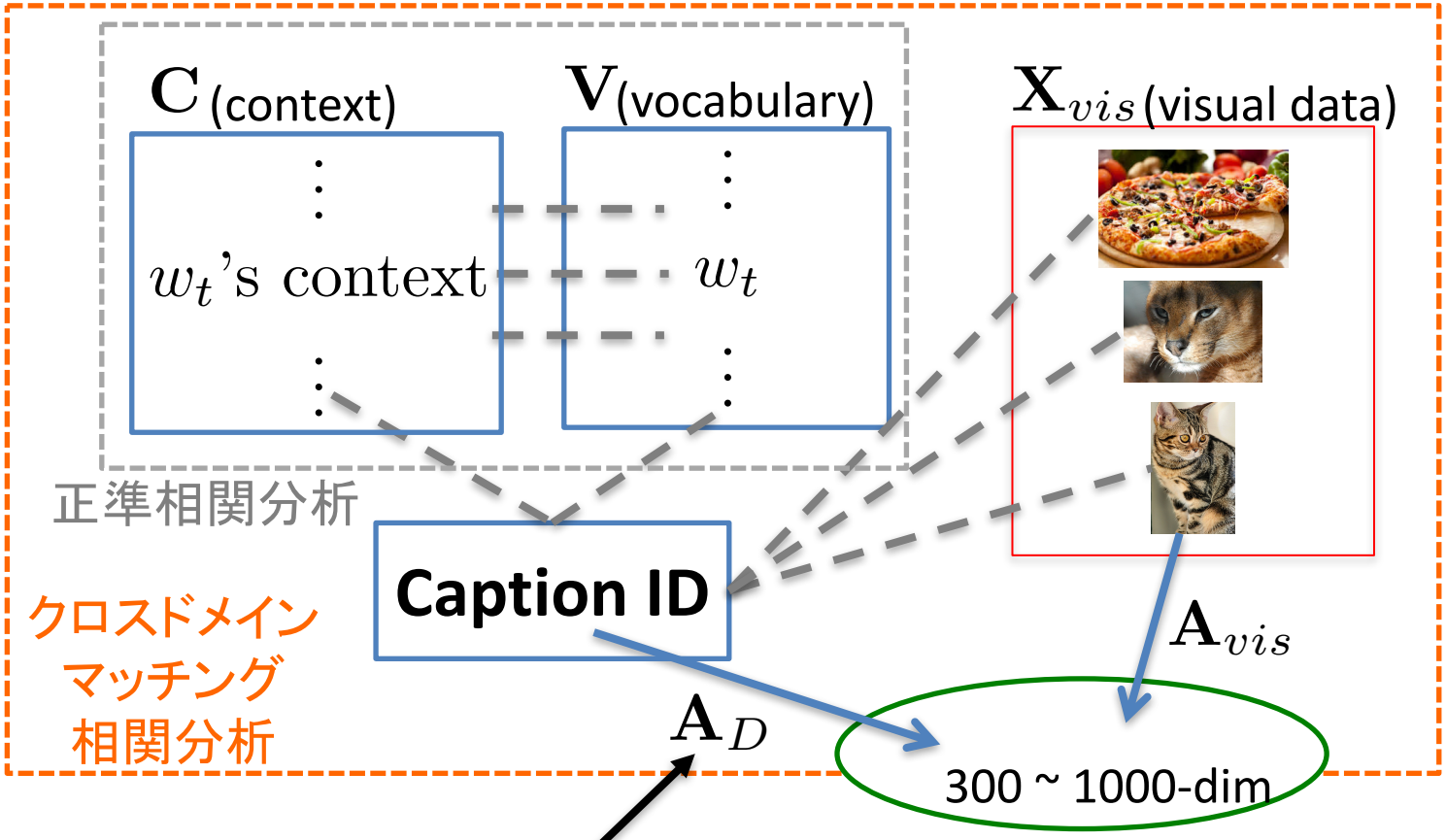


画像情報を反映したキャプションのベクトル表現

※学生優秀プレゼンテーション賞対象

D1-53：クロスドメインマッチング相関分析を用いた画像とキャプションの同時埋め込み

京都大学大学院情報学研究科 / 理研AIP
福井一輝，下平英寿



画像情報を反映したキャプションのベクトル表現

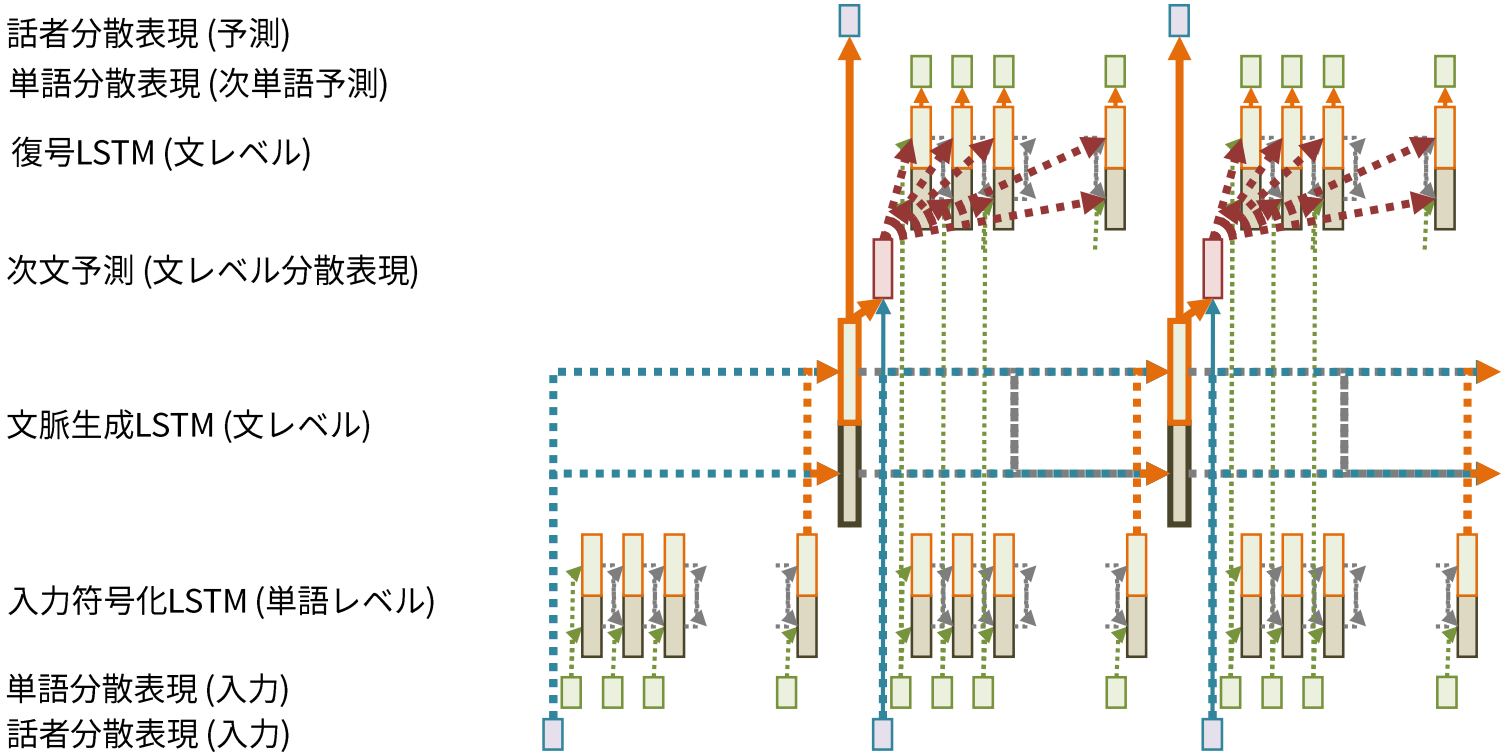
※学生優秀プレゼンテーション賞対象

Web 小説を学習に用いた二段階 LSTM による台本形式小説生成

D1-54 鈴木 惇 (東京大学・情報理工)

学生優秀プレゼンテーション賞対象

- Web上の台本形式小説 (話者ラベル付き会話文)
  - のび太「ドラえもーん」
  - ドラえもん「また怒られたのか、のび太くん」
  - のび太「だってジャイアンが…」
- 自動生成→提案手法: 二段階結合LSTM相互会話生成器 (TULIP)

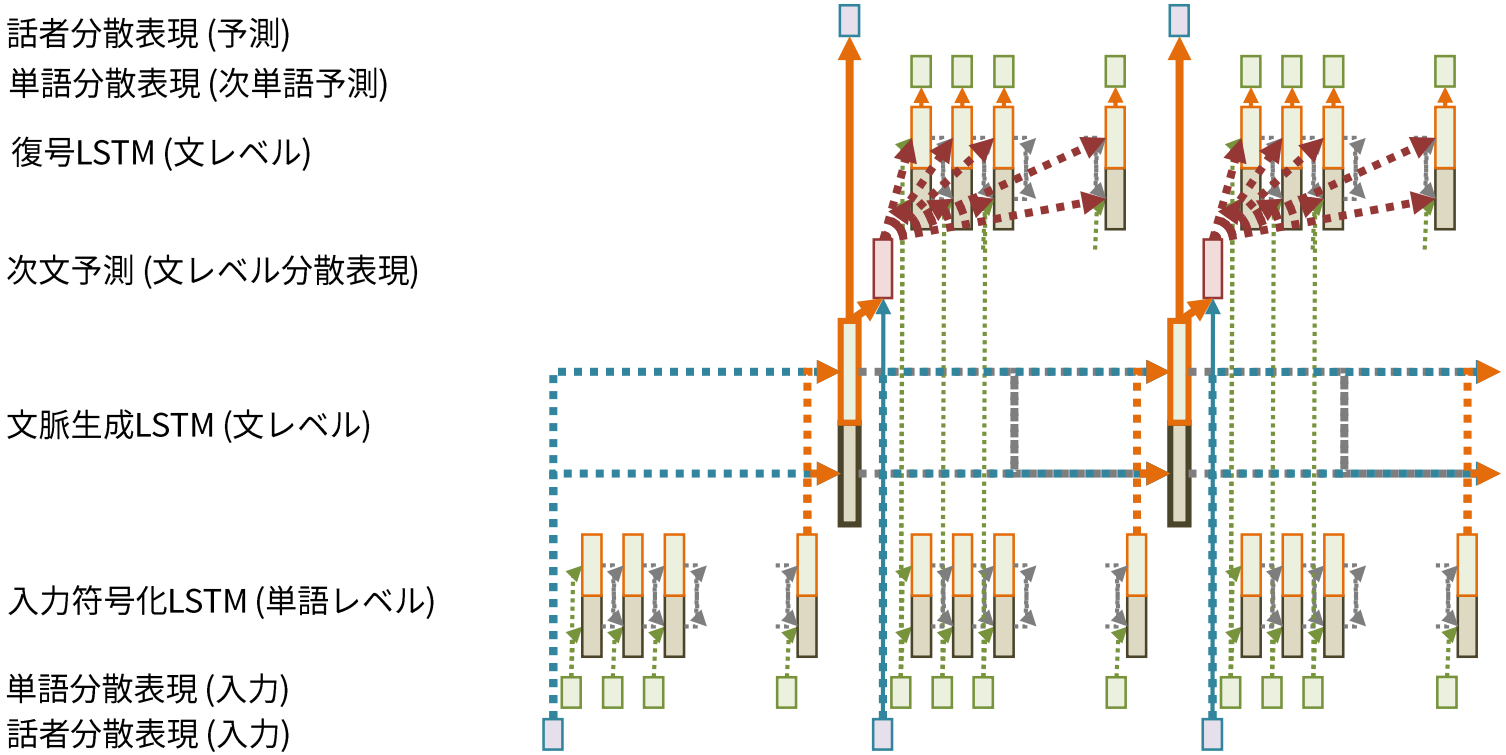


Web 小説を学習に用いた二段階 LSTM による台本形式小説生成

D1-54 鈴木 惇 (東京大学・情報理工)

学生優秀プレゼンテーション賞対象

- Web上の台本形式小説 (話者ラベル付き会話文)
  - のび太「ドラえもーん」
  - ドラえもん「また怒られたのか、のび太くん」
  - のび太「だってジャイアンが…」
- 自動生成→提案手法: 二段階結合LSTM相互会話生成器 (TULIP)



D1-55

密度比推定を用いた条件付き相互情報量の近似

高品佑也（早大）・中谷秀洋（サイボウズ・ラボ）・井上真郷（早大）

Markov確率場の構造推定

- 辺の有無が条件付き独立性に対応
- 条件付き独立性は条件付き相互情報量に対応

$$I\left(X_i, X_j \mid \boldsymbol{X}_{\setminus\{i,j\}}\right) = 0 \Leftrightarrow X_i \perp\!\!\!\perp X_j \mid \boldsymbol{X}_{\setminus\{i,j\}}$$

⇒条件付き相互情報量が分かればグラフ構造が分かる

本研究

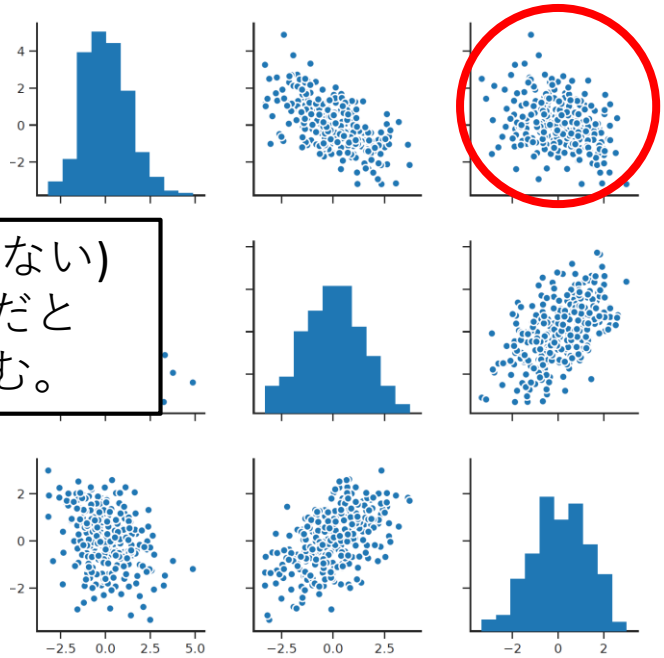
- **すべての変数が連続な場合**に密度比推定により**条件付き相互情報量**を近似
- 密度比推定の基底が無相関Gauss分布の場合に正規化定数の計算を高速化 $O\left(N^3\right) \to O(N)$

D1-55

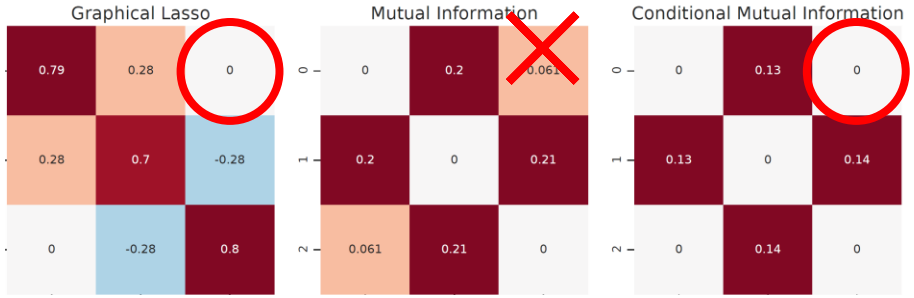
密度比推定を用いた 条件付き相互情報量の近似

高品佑也（早大）・中谷秀洋（サイボウズ・ラボ）・井上真郷（早大）

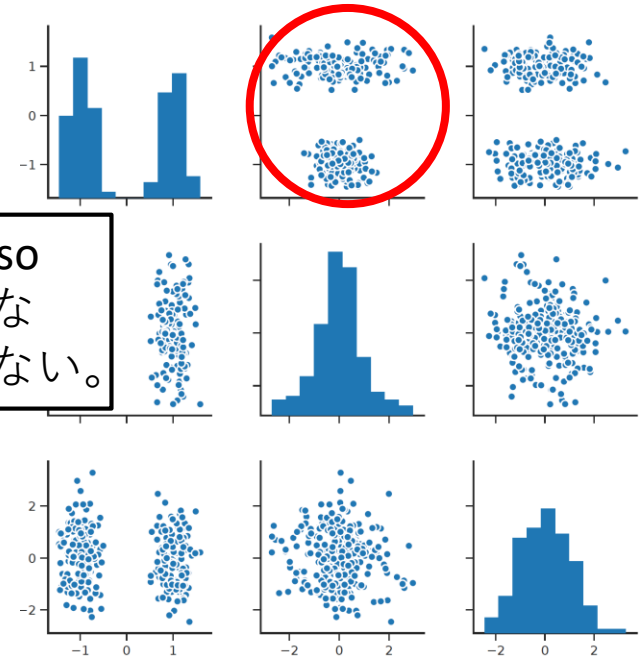
実験 1: 多変量Gauss分布



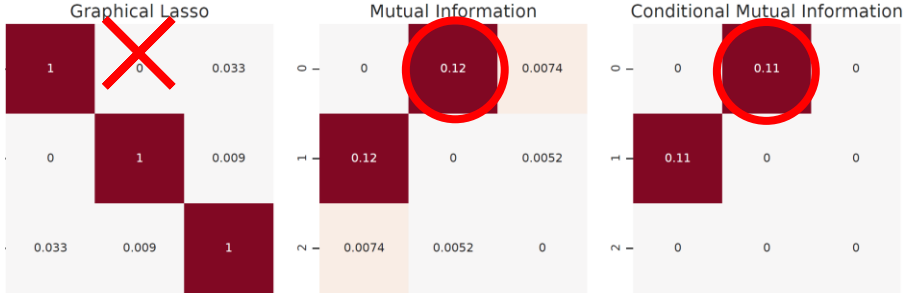
(条件付きでない)
相互情報量だと
偽相関を含む。



実験 2: 混合Gauss分布



Graphical lasso
だと非線形な
関係が取れない。



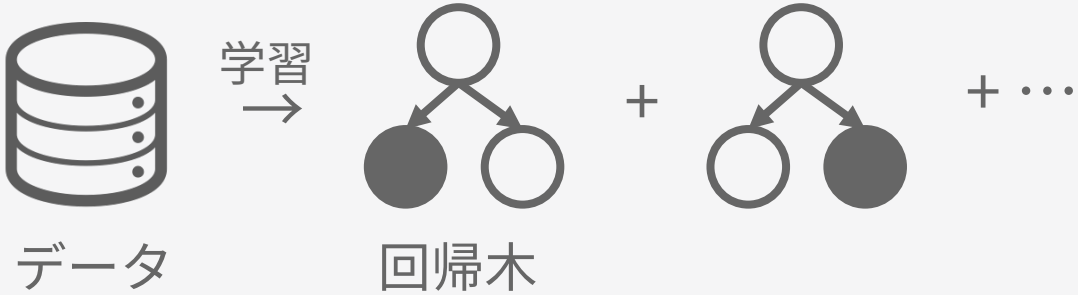
記述長最小原理に基づく ブースティング木学習の早期停止基準

D1-56

花田悠介(東京大学, CREST), 山西健司(東京大学, CREST)

学生優秀プレゼンテーション賞対象

Boosted Treesの学習の早期停止



木が多すぎると…
過学習、計算コスト大
木が少なすぎると…
未学習

適切な早期停止基準が必要

提案手法

データとモデルの記述長が最小となるような木の数で学習を停止

- CVより計算コスト低
- 優れた汎化性能

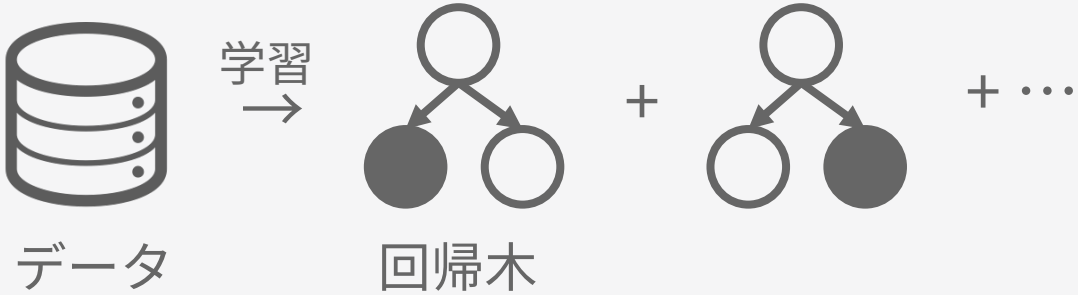
記述長最小原理に基づく ブースティング木学習の早期停止基準

D1-56

花田悠介(東京大学, CREST), 山西健司(東京大学, CREST)

学生優秀プレゼンテーション賞対象

Boosted Treesの学習の早期停止



木が多すぎると…
過学習、計算コスト大
木が少なすぎると…
未学習

適切な早期停止基準が必要

提案手法

データとモデルの記述長が最小となるような木の数で学習を停止

- CVより計算コスト低
- 優れた汎化性能

オンラインジャッジサイトのデータセットのプログラム自動生成への応用

D1-57

P.C. Nagesh (メルボルン大), 江原 遥 (産総研), 徳永 拓之（スマートニュース）, 鶴岡 慶雅（東大）, 小田 悠介（NAIST/NICT）, 渡部 有隆（会津大）

プログラム自動生成とは:

入力: 自然言語　出力: プログラム

課題:

1. 訓練データとして必要な（自然言語, プログラム）対形式のデータセットが少ない。作成も高コスト。
2. プログラムに自然言語による説明を加えたデータは多いが、本来の流れである、自然言語による説明からプログラムを書くことで作成した対形式のデータセットはさらに少ない

本研究の提案

キーアイデア：

1. プログラミングコンテストのオンラインジャッジサイトでは、自然言語による問題文に対して、その問題を解くコードを大勢のユーザが投稿し、コードが日々蓄積されている
2. しかも、問題文を人間が読みプログラムを作成する順方向のデータになっている

→日本の代表的なオンラインジャッジサイトAizu Online Judgeの問題文・投稿コードを整備・翻訳して、データセットを作成中！

https://aistairc.github.io/plu/ にて詳細を公開予定

予備実験：（問題文,コード）ペアに対して、コードで問題が解けるかを判定するタスクの精度で性能評価。

D1-58

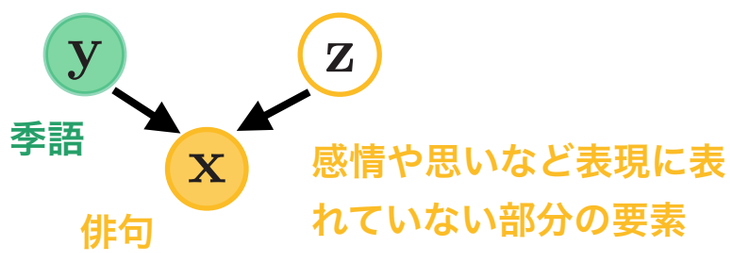
学生優秀プレゼンテーション対象

条件付き変分オートエンコーダを用いた俳句生成

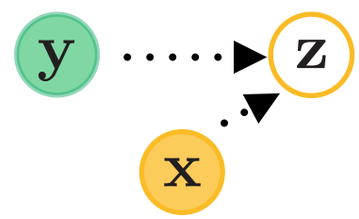
川島寛乃　河野慎　米澤拓郎　中澤仁（慶應義塾大学）

研究方法 俳句の言語モデルの推定

生成過程



推論過程

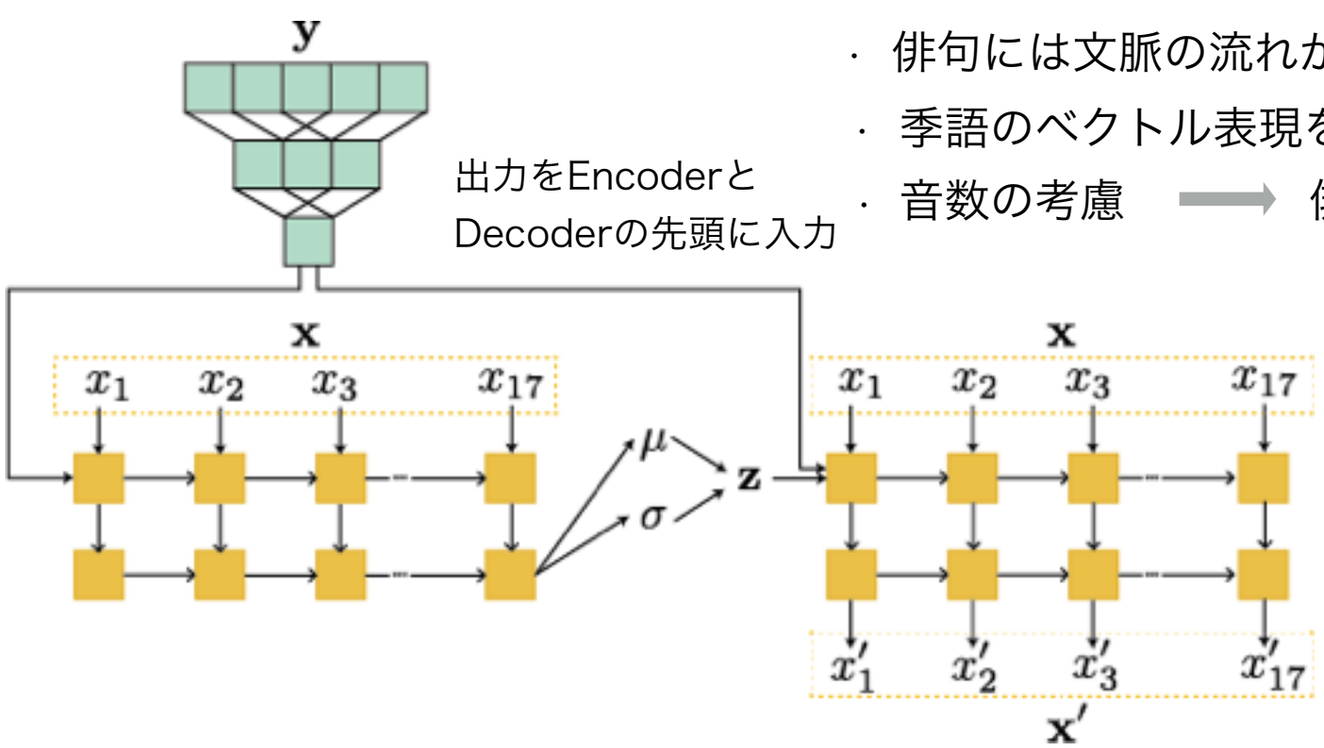


$$p(\mathbf{x},\mathbf{y}) = \int p(\mathbf{x}|\mathbf{y},\mathbf{z})p(\mathbf{y})p(\mathbf{z})d\mathbf{z}$$

$$= \int p(\mathbf{y})p(\mathbf{z})\prod_{t=1}^{T} p(\mathbf{x}_t|\mathbf{x}_{<t},\mathbf{y},\mathbf{z})d\mathbf{z}$$

$p(\mathbf{z})$ …ガウス分布　$p(\mathbf{x}|\mathbf{y},\mathbf{z})$ …マルチヌーイ分布

アプローチ 条件付き変分オートエンコーダの適用



- 俳句には文脈の流れがある ⟶ EncoderにLSTMを採用
- 季語のベクトル表現を獲得 ⟶ CNNを採用
- 音数の考慮 ⟶ 俳句と季語をひらがなで扱う

実験

データセット：Webから収集した俳句(168819句)

季語	生成した俳句
虫時雨	むししぐれ ふうのまつまの なつしばい
浮寝鳥	たいたいの ひとつきてゐる うきねとり
冬ざれ	ふゆざれや いっしょうたるる やまぶんろ

D1-58

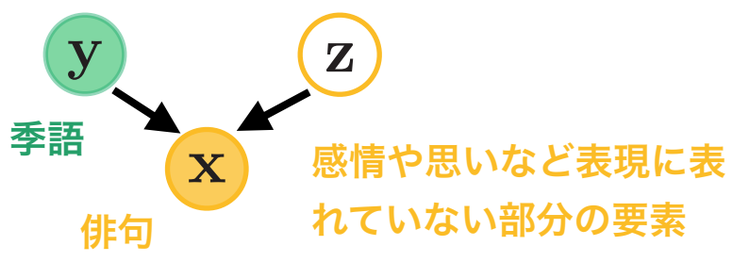
学生優秀プレゼンテーション対象

条件付き変分オートエンコーダを用いた俳句生成

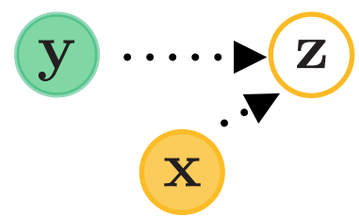
川島寛乃　河野慎　米澤拓郎　中澤仁（慶應義塾大学）

研究方法 俳句の言語モデルの推定

生成過程



推論過程

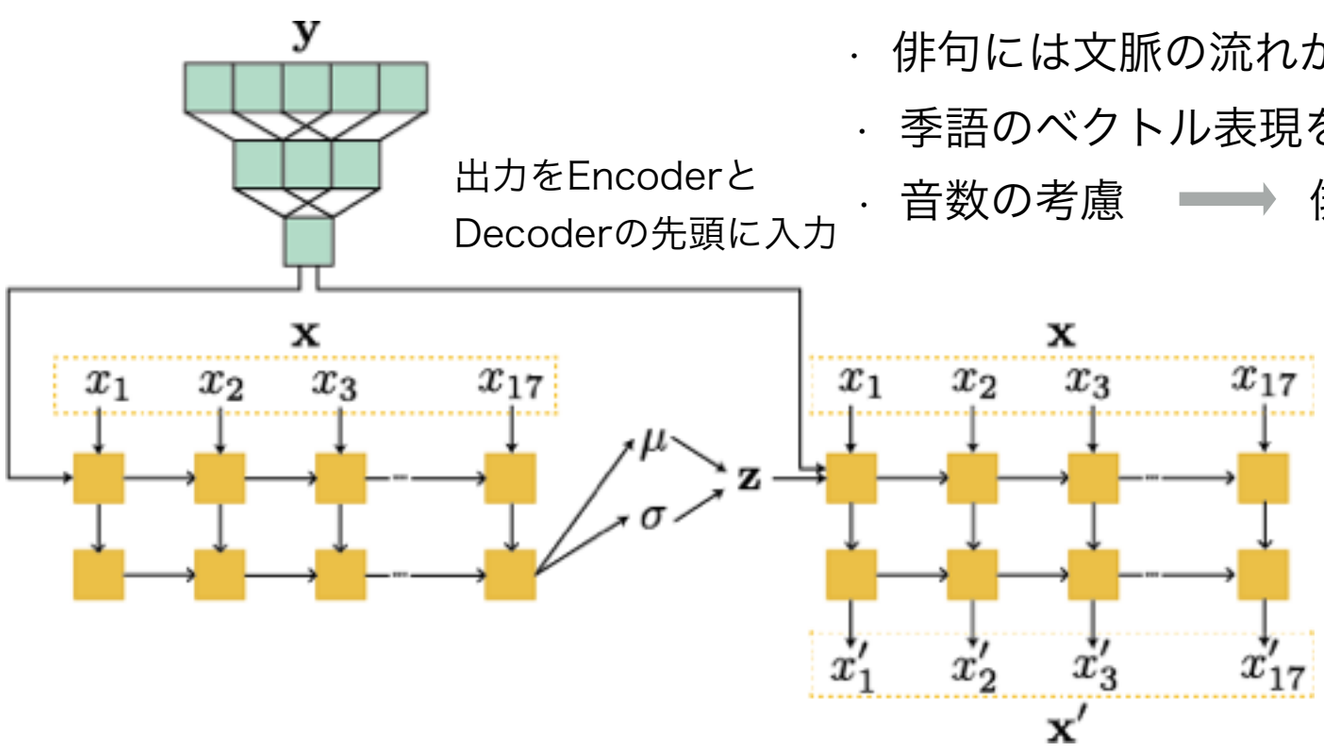


$$p(\mathbf{x},\mathbf{y}) = \int p(\mathbf{x}|\mathbf{y},\mathbf{z})p(\mathbf{y})p(\mathbf{z})d\mathbf{z}$$

$$= \int p(\mathbf{y})p(\mathbf{z})\prod_{t=1}^{T} p(\mathbf{x}_t|\mathbf{x}_{<t},\mathbf{y},\mathbf{z})d\mathbf{z}$$

$p(\mathbf{z})$ …ガウス分布　$p(\mathbf{x}|\mathbf{y},\mathbf{z})$ …マルチヌーイ分布

アプローチ 条件付き変分オートエンコーダの適用



- 俳句には文脈の流れがある ⟶ EncoderにLSTMを採用
- 季語のベクトル表現を獲得 ⟶ CNNを採用
- 音数の考慮 ⟶ 俳句と季語をひらがなで扱う

実験

データセット：Webから収集した俳句(168819句)

季語	生成した俳句
虫時雨	むししぐれ ふうのまつまの なつしばい
浮寝鳥	たいたいの ひとつきてゐる うきねとり
冬ざれ	ふゆざれや いっしょうたるる やまぶんろ

D1-59: 交差検証による正常データ選別の自動化

HITACHI
Inspire the Next

鈴木英明，藤原淳輔，石田響子，河野敏明（(株)日立製作所 研究開発グループ）

■課題：学習させるデータの適切なラベリング

- 様々な機器や設備の状態は，正常・故障の二元論で定義できるものではなく，正常から少し外れた状態や故障の前兆段階まで様々な状態が存在する
- 機械学習を用いて時系列データに基づき異常検知を行う場合，「正常データ」を学習し再構成誤差を指標とする方法では，「正常データ」の与え方そのものが重要であり，機器の「正常」を規定することは実際には難しい

■アプローチ：交差検証によるスコアリング

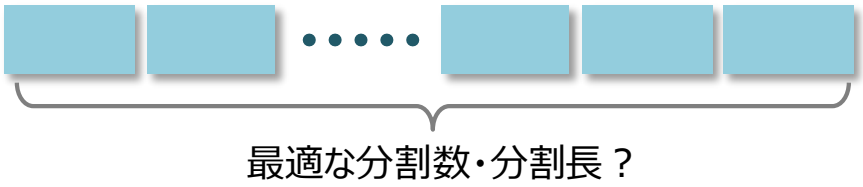
- 交差検証としては，K-分割やleave-one-outが用いられるが，分割の仕方や学習と検証データの比率について調べた例はあまり見当たらない
- 本研究では，分割単位などについて実データを用いて比較検討を行い，最適な「学習データ」の選択について考察する

© Hitachi, Ltd. 2017. All rights reserved.

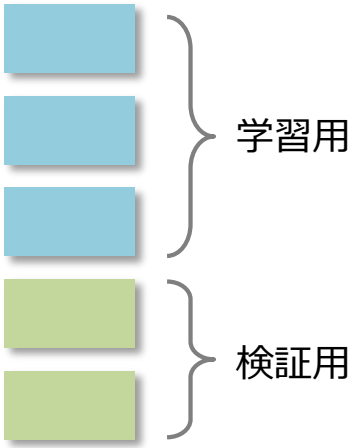
D1-59: 交差検証による正常データ選別の自動化

HITACHI
Inspire the Next

■分割単位
・分割数あるいは分割長による違いはあるのか？



■学習・検証データの比率
・学習用と検証用のデータの比率はどのくらいが適正か？
・比率の最適値があるとしたら，何に依存しているのか？



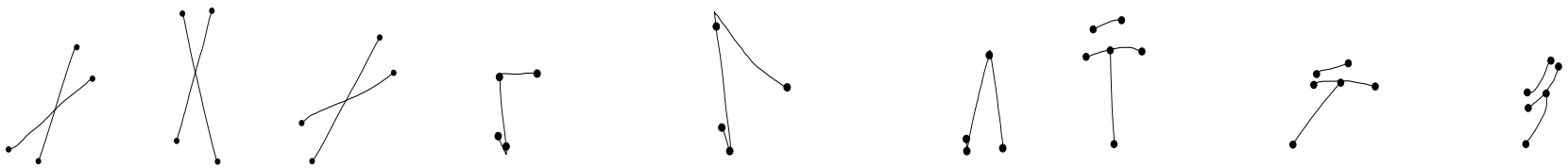
■スコアリングに基づく「正常データ」の選択
・交差検証を援用し，再構成誤差をスコアに使う場合，自動的にその閾値を決定するには基準をどのように定めるべきか？

© Hitachi, Ltd. 2017. All rights reserved.

D1-60: フルプロクルステス距離の拡張とその距離計量学習

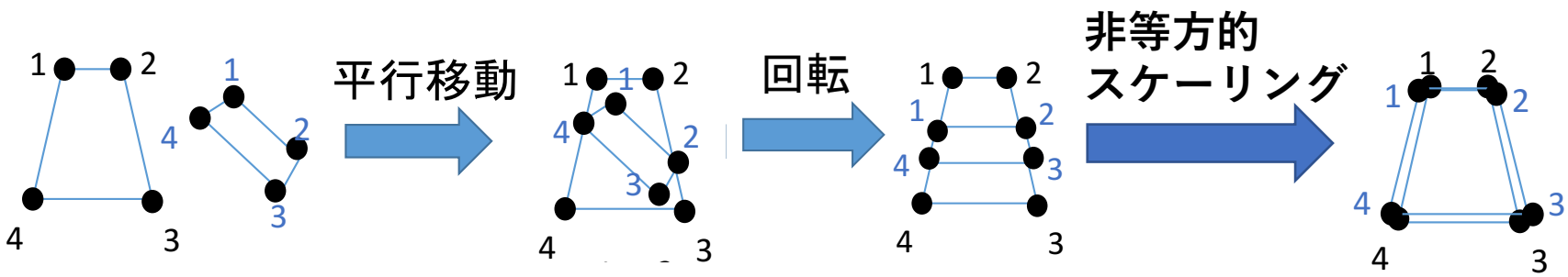
岡本　司，岩田一貴，末松伸朗 (広島市立大学)

- ランドマークに基づく形状間の距離を用いた形状検索について発表する．



線描画形状の例(点はランドマーク)

- フルプロクルステス距離では，非相似に歪んだ形状を扱うことができないという問題がある．そこで，距離計量学習を用いて，非等方的スケーリングを学習することによって，フルプロクルステス距離を拡張し，この問題を解決する．



形状データセットに対しての検索率(%)

距離	データセット		
	X,平方根,Tの線描画	bluetooth,郵便局,魚の線描画	ETHZ(Ferrari, 2009)
部分プロクルステス距離	81.0	99.4	97.6
リーマン距離	81.0	99.4	97.6
フルプロクルステス距離	81.0	99.4	97.6
拡張フルプロクルステス距離	95.1	100	98.3

- 検索率はブルズ・アイ評価(Bulls-eye test)によって計算した.
- 各データセットの線描画・輪郭の形状はそれぞれ4,6,6点のランドマークで表現した.

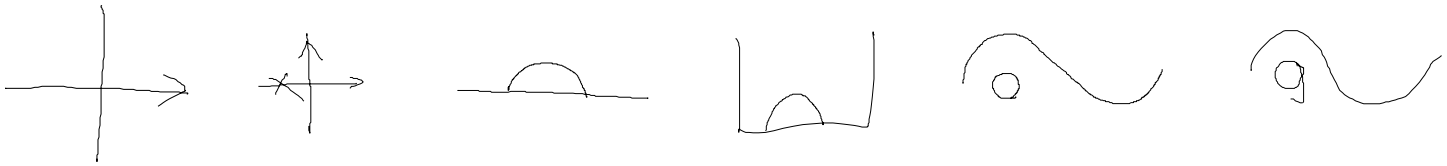
従来の形状距離よりも検索率が高い

V. Ferrari, “ETHZ shape classes,” http://www.vision.ee.ethz.ch/datasets/index.en.html, Sep. 2009.

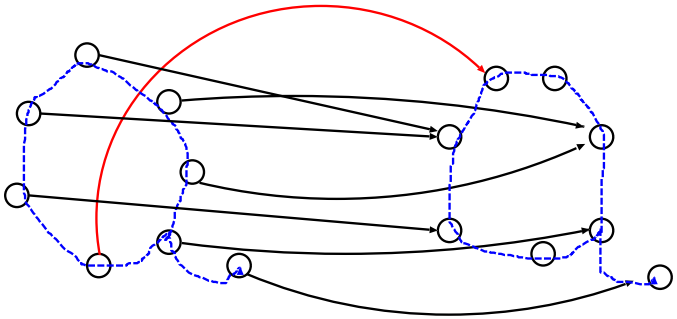
D1-61：近傍標本点の対応を考慮した形状間の対応づけとその計算コストの削減

山本大貴，岩田一貴，末松伸朗（広島市立大学）

- 向きのない（書き順に依存しない）線画形状の間の距離を使った形状検索について発表する．



- 距離は形状間の対応づけ（整合）に依存するが，その対応づけがいくつか間違った場合に，それを自動で訂正するように近傍標本点の保存を考慮した手法を提案する．



- また，提案した手法を分岐限定法による近似を使って高速に実行するための方法について検討する．

手法	近傍標本点数		
	1	2	3
従来の手法	67.075 (2分31秒)	67.075 (2分31秒)	67.075 (2分31秒)
最近傍点基準 (分岐限定法による近似あり)	**67.625** (4分57秒)	**67.55** (4分56秒)	**67.55** (4分58秒)
最近傍点基準 (近似なし)	67.625 （約11日）	67.55 （約11日）	67.55 （約11日）

・括弧内は実験の実行時間.

➡ 近傍標本点の保存を考慮しないときと比べて，最近傍点基準によってそれを考慮した方が検索率が高い.

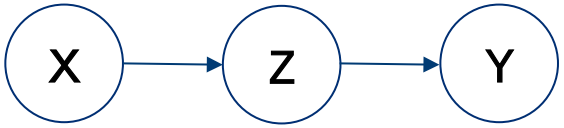
➡ 括弧内より分岐限定法による近似はかなり有効.

D1-62: ソフトラベル判別

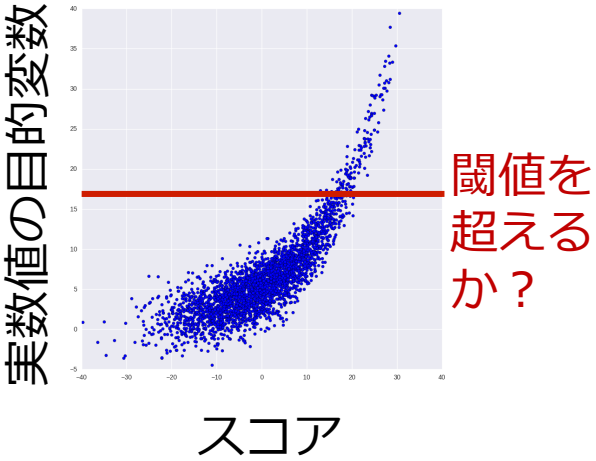
谷本啓 (NEC / 京大) 、山田聡 (NEC)

問題：実数値の目的変数が ある閾値を超えるかどうか判別

- causal-setting の判別で、中間値Zが得られる場合
- 例：商品が100個以上売れそうなら陳列する



商品の説明変数　売上数　100個以上売れるか



一般的な手法：

- 実数値を0/1に縮約して直接判別：× 実数値の情報が失われ、推定分散が増加
- 回帰モデルの予測値+閾値で判別：× 間接的であるため推定バイアスが増加

提案手法：実数値の情報を反映したソフトラベルを学習

✓ 正例が少数でも安定した学習が可能
✓ 出力は判別モデルと同じなので解釈性の損失なし
✓ 実装容易かつGPやランキング学習を用いるより高速

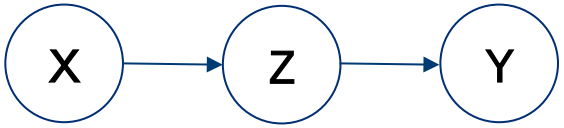
© NEC Corporation 2017

D1-62: ソフトラベル判別

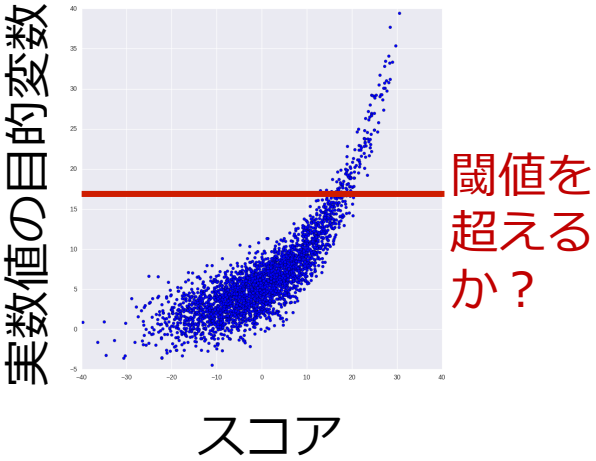
谷本啓 (NEC / 京大) 、山田聡 (NEC)

問題：実数値の目的変数が ある閾値を超えるかどうか判別

- causal-setting の判別で、中間値Zが得られる場合
- 例：商品が100個以上売れそうなら陳列する



商品の説明変数　売上数　100個以上売れるか



一般的な手法：

- 実数値を0/1に縮約して直接判別：× 実数値の情報が失われ、推定分散が増加
- 回帰モデルの予測値+閾値で判別：× 間接的であるため推定バイアスが増加

提案手法：実数値の情報を反映したソフトラベルを学習

✓ 正例が少数でも安定した学習が可能
✓ 出力は判別モデルと同じなので解釈性の損失なし
✓ 実装容易かつGPやランキング学習を用いるより高速

© NEC Corporation 2017