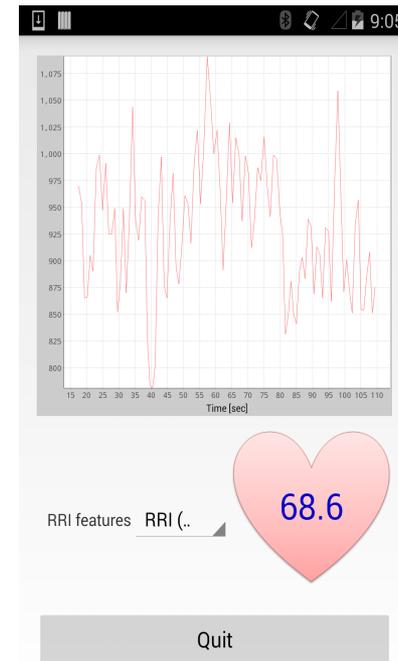
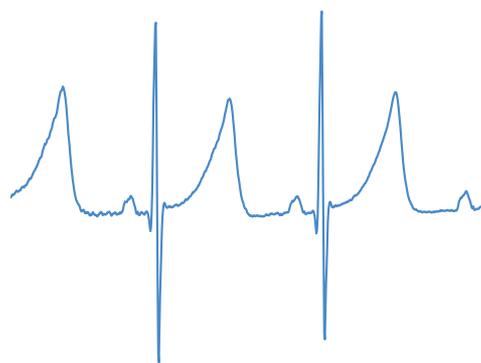


心拍変動解析と SVM に基づいた 睡眠時無呼吸症候群 スクリーニングシステムの開発

○仲山千佳夫 京都大学ヒューマンシステム論分野研究室



D1-31にて実機を用いたデモを行います

※学生優秀プレゼンテーション賞対象

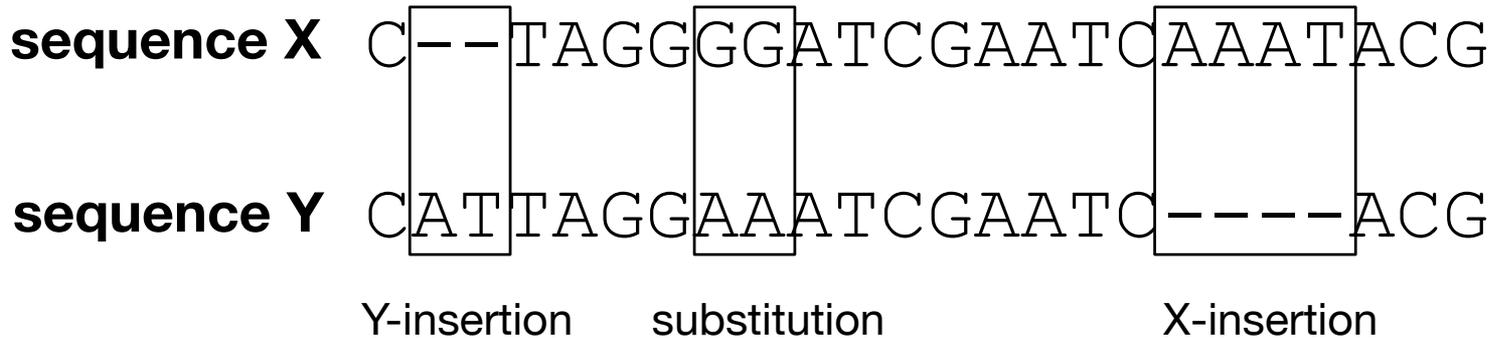


Model Selection on Pairwise Hidden Markov Models using Factorized Information Criterion

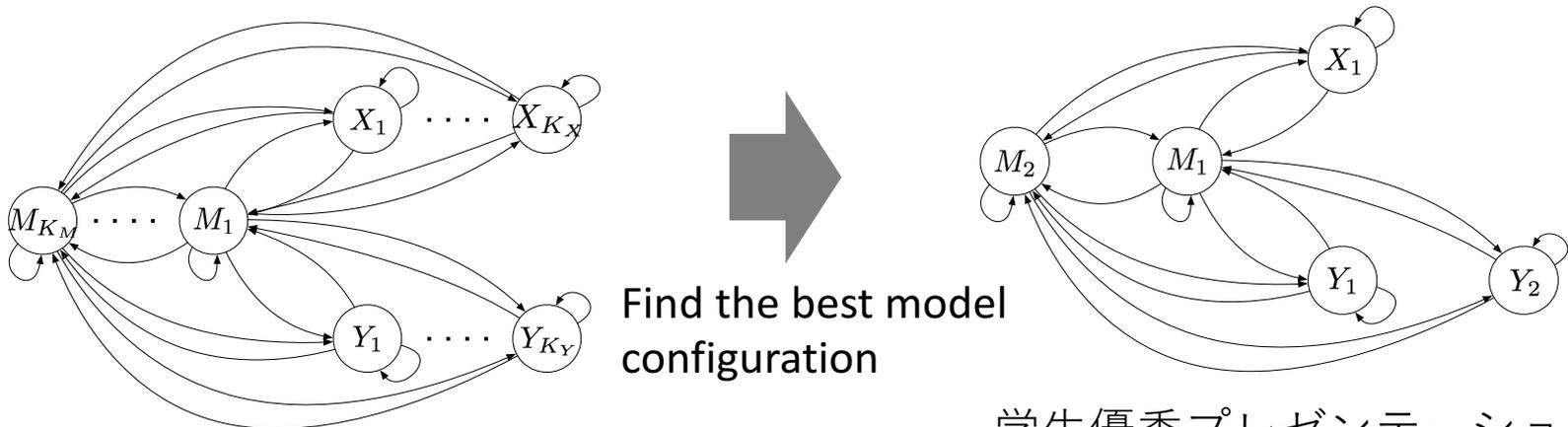
D1-32

Taikai Takeda¹, Michiaki Hamada^{1,2} ¹Waseda University, ²AIST-Waseda CBBDOIL

Question - What is the “best” model for biological sequence alignment?



How? - Maximization of Factorized Information Criterion (FIC) on Pairwise Hidden Markov Models (PHMM)



学生優秀プレゼンテーション賞対象

D1-33 : Bayes factorを用いた 大規模ベイジアンネットワークの構築

名取和樹, 宇都雅輝, 植野真臣 (電気通信大学 大学院情報理工学研究科)

ベイジアンネットワークの構造学習はNP困難であり、
既存の厳密解探索手法では、**最先端手法でも60変数程度の学習が限界**
一方、条件付き独立性テストに基づく制約付き構造探索法は
計算効率が良いが、**漸近一致性を持たず、真の構造学習の保証がない**

本研究の提案

漸近一致性を有するBayes factorを用いた条件付き独立性テストを開発
本手法を組み込んだ新たな制約付き構造探索法を提案

⇒ 提案手法により、漸近一致性を有したまま計算量を**画期的に削減**

実験により、**1000変数以上の構造学習を実現**できることを示す

マージン最大化にもとづく確率的ブロックモデル

小山田 昌史, 中台 慎二 (NEC)

[従来] 確率的ブロックモデル (SBM) [Wang+; JASA'87]

- ・ネットワークのリンク生成過程をパラメトリックに生成モデルでモデル化
- ・ネットワークから「静的なコミュニティ構造」が発見可能

[提案] Max-Margin SBM

- ・リンク生成過程をノンパラメトリックに判別モデル (SVM) でモデル化
- ・ネットワークから「状況に応じたコミュニティ構造」が発見可能



D1-35

比較バンディット問題における コーブランド勝者の推薦

- K個のアーム（オプション）
- RQ: 何度の一対比較で、最も良いアームを発見できるか？
- **コンドルセ勝者**の発見 [小宮山 本多 鹿島 中川, IBISML沖縄2015] を一般化：コンドルセ勝者がいない場合も **コーブランド勝者**を発見可能

コーブランド勝者

	1	2	3	4	5	L_i
1	0.5	0.7	0.7	0.7	0.4	1
2	0.3	0.5	0.7	0.4	0.6	2
3	0.3	0.3	0.5	0.6	0.6	2
4	0.3	0.6	0.4	0.5	0.4	3
5	0.6	0.4	0.4	0.6	0.5	2

小宮山純平 本多淳也 中川裕志（東京大学）

カーネル法を用いた 従属性の高い部分木ペアの教師なし推定

横井祥*, 持橋大地**, 岡崎直観*, 乾健太郎* *東北大 **統数研

背景

コーパスから知識 (e.g., $P \rightarrow Q$) を取り出す際, 自然言語文に含まれるノイズを取り除く必要がある

解くべき問題

自然言語処理でよく見られる問題設定

入力: 木 (=文) のペアの集合 $\mathcal{D} = \{(x_i, y_i)\}_i$

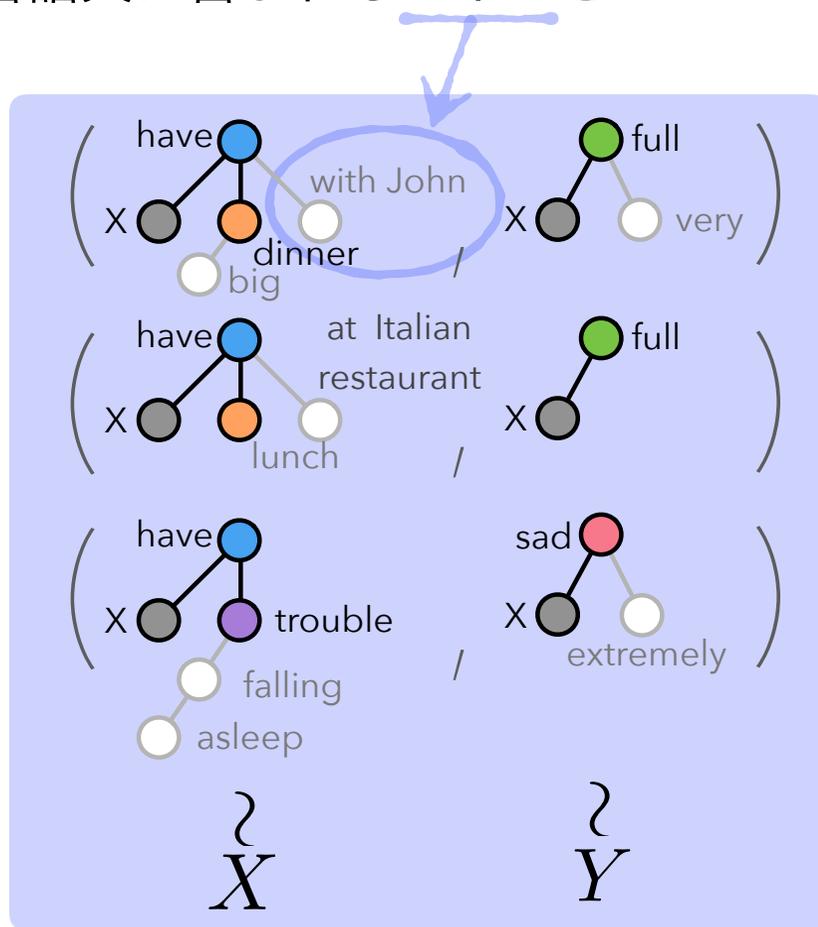
↓ 枝刈り

出力: 根付き部分木のペアの集合 $\mathcal{D}' = \{(x'_i, y'_i)\}_i$

ペアをペアたらしめている“コア”の部分
(=自己相互情報量の高い部分構造ペア)
を取り出したい

提案手法

カーネルを用いた独立性尺度により
高次元空間で確率変数対の従属性を最大化



[D1-37] 確率的決定則について

NTT コミュニケーション科学基礎研究所 村松 純

NTT 未来ねっと研究所 三宅茂樹

ベイズ決定則

- 観測データ Y からこれと相関のあるラベル X の値を判定する問題を考える.
- 事後確率最大決定則は誤り確率を最小にする.

確率的決定則

- 判定を確率的(ランダム)に行うことを考える.
- 事後確率分布を用いた確率的決定則の誤り確率と事後確率最大決定則の誤り確率の関係を議論する.
- 情報圧縮の復号アルゴリズムへの応用について触れる.

[おことわり] 本ポスターの完全版を来月開催されるの情報理論とその応用シンポジウム (SITA2016) にて発表する予定です (予稿あり).

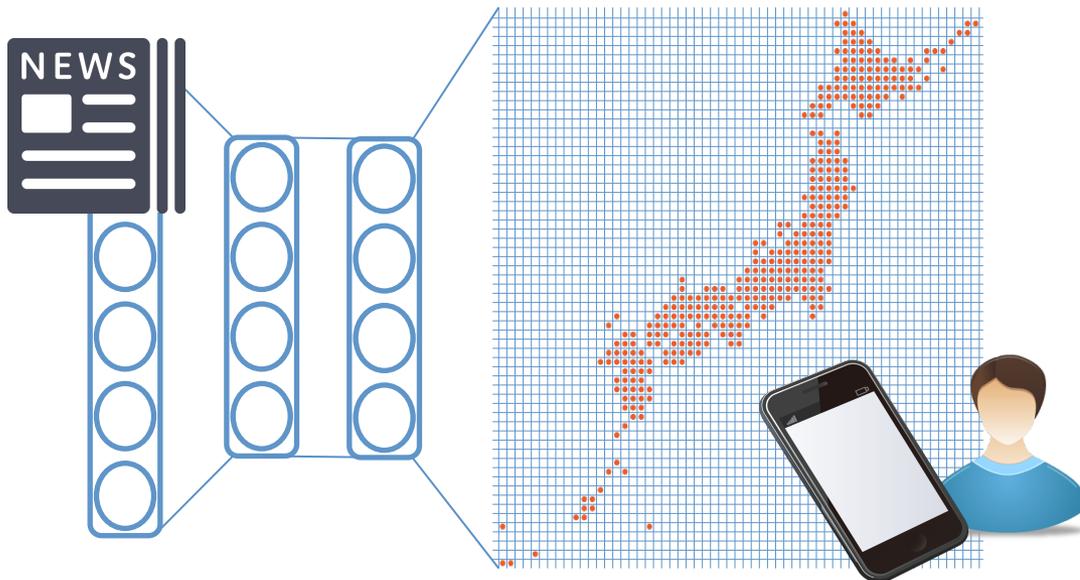
D1-39

クリックフィードバックを用いた 記事の地域性推定モデルの構築

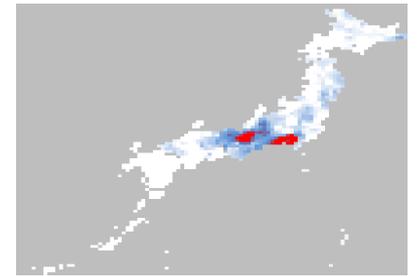
YAHOO!
JAPAN

大倉 俊平 (ヤフー株式会社)

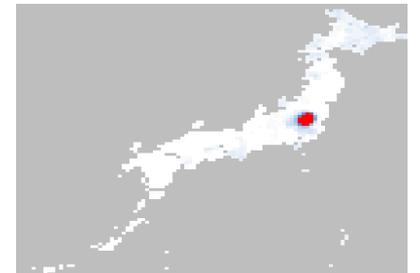
- 記事レコメンドのクリックフィードバックから、地域性の強い語とその関連エリアを学習
- 地名などの辞書を用いず**に、「単語/記事」→「エリア」の対応付けを獲得



東海道
新幹線



作新



条件付き確率場を用いた

自然言語文章からの上位下位関係抽出

D1-40

平松淳, 若林啓 (筑波大学) * 学生優秀プレゼンテーション賞対象



目的

自然言語文章からの
上位下位関係の自動抽出

isA(富士山, 山)

isA(Ruby, 言語)

提案手法

上位下位関係抽出を
系列ラベリング問題とみなす

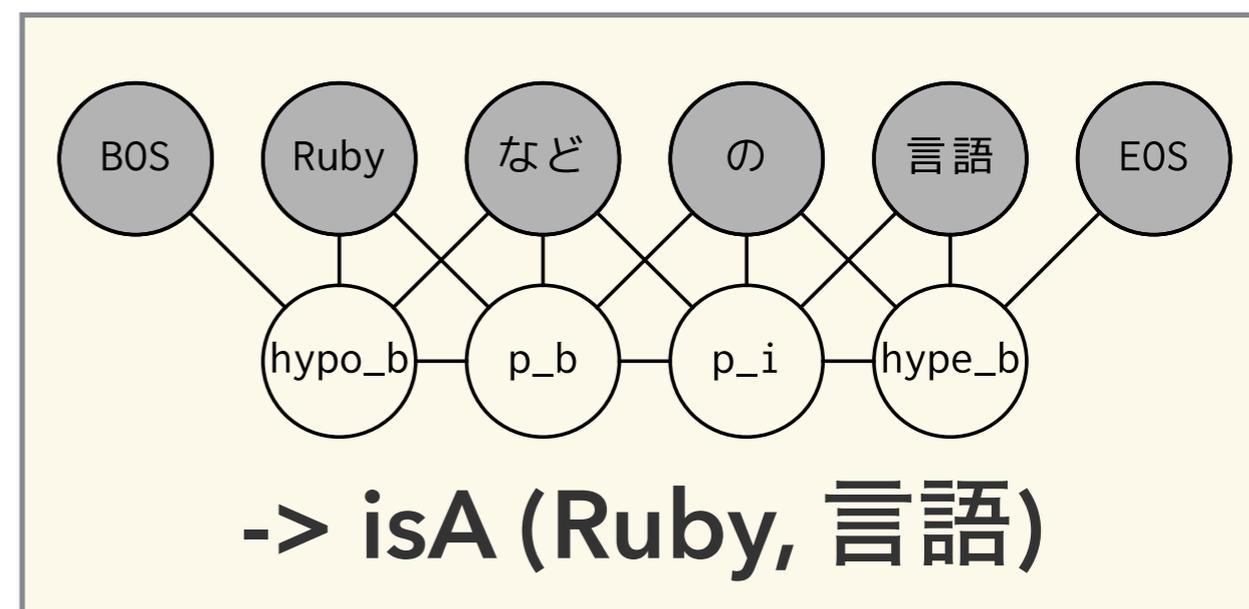
関連研究

ルールベース [Hearst+, 1992]

(精度が低い)

頻度ベースの反復学習 [Wu+, 2012]

(大規模なコーパスが必要)



D1 光干渉断層計による網膜層厚

-41 からの緑内障患者の視野推定について

※ 本研究はJST-CRESTの一部として行われた

○上坂俊允, 森野佳生, 村田博史, 朝岡亮, 山西健司(東大, CREST)

目的

網膜層厚からの各点視野感度推定

背景

- ・ 視野感度の測定は臨床現場にとって負担
- ・ 網膜情報を用いた視野上各点感度推定法がまだ無い
- ・ 視野のみ, 網膜のみを計測した非教師データの存在

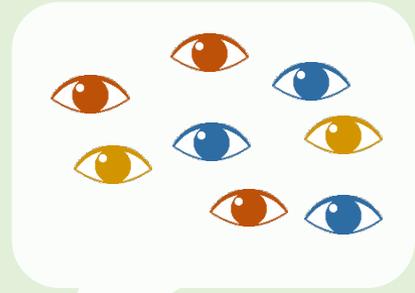
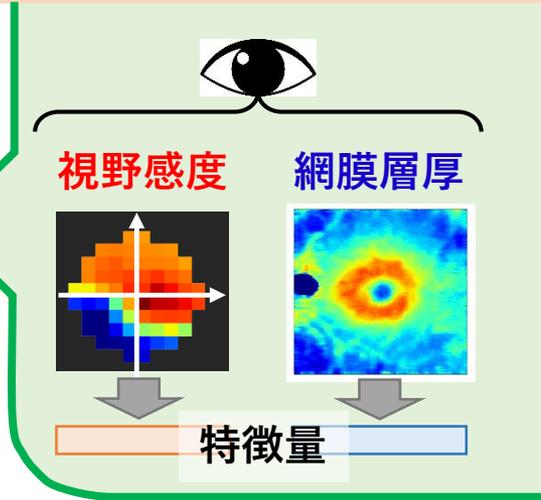
アプローチ

非教師データをも用いて各ドメインで特徴量を抽出

緑内障病状のヘテロ性を考慮した教師データの選別

異なるドメインの特徴量をつなぐ潜在変数モデル学習

特徴量から視野データ復元
各視野点上の予測を実現



傾向の似た眼を選別

共通因子

網膜層厚
特徴量

視野特徴量

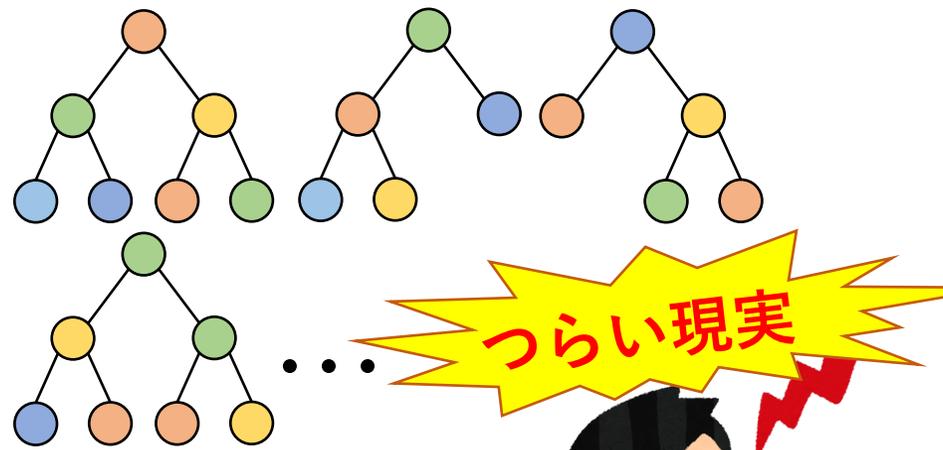
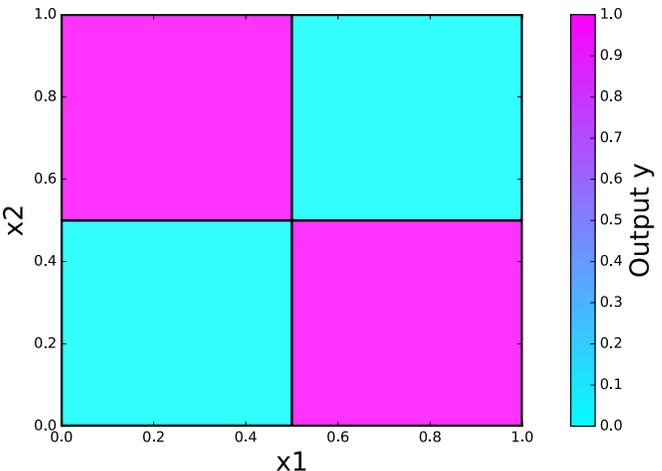
予測対象(欠損)

アンサンブル木モデル解釈のためのモデル簡略化法

原聡 (NII/河原林巨大グラフプロジェクト)、林浩平 (産総研)

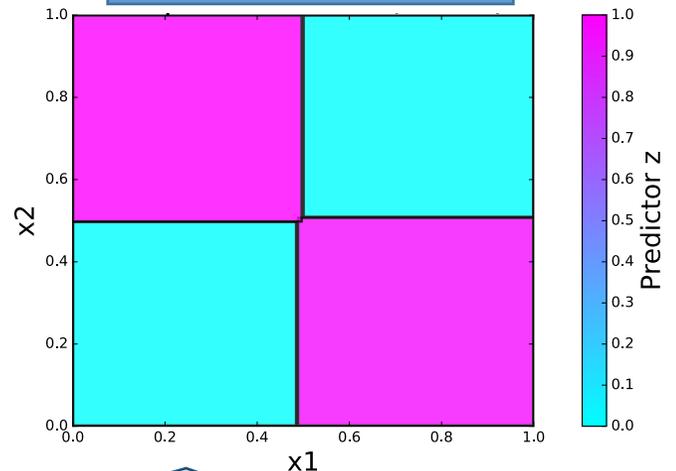
- アンサンブル木モデル：ランダムフォレスト、Gradient Boosted Treesなど
 - ・ 長所：予測精度が高い。様々なデータに適用できる。
 - ・ 短所：学習したモデルの解釈性が低い。

GitHub: sato9hara/defragTrees



つらい現実

決定木が多すぎて、どんなモデルが学習されたのかわからない！！



このモデルは4つのルールで(近似的に)記述できるよ！各ルールはこんなだよ！

D1-43

非負制約付きスパースコーディングを用いた チームスポーツフォーメーション特徴量の自動抽出

阿部 俊樹, 山本 莉沙, 中田 洋平 (明治大学)



- 目的：チームスポーツに対するフォーメーション解析の高度化
- 方法：非負制約付きスパースコーディングによる特徴量の自動抽出

① 変換フェーズ：
選手位置情報から
選手密度情報を算出

選手位置情報 選手密度情報

② 学習フェーズ：
選手密度情報を用いた辞書学習

基底選手密度の辞書
(フォーメーション部品)

③ 算出フェーズ：
得られた辞書を用いた特徴量
(係数) の算出

基底選手密度の係数
(特徴量)

非負制約付きスパースコーディング

提案手法

浦和 横浜

ボール奪取 ゴール付近に到着

浦和 横浜

前フレームと現在フレーム
の特徴量差の2ノルム

フレーム番号

実験結果

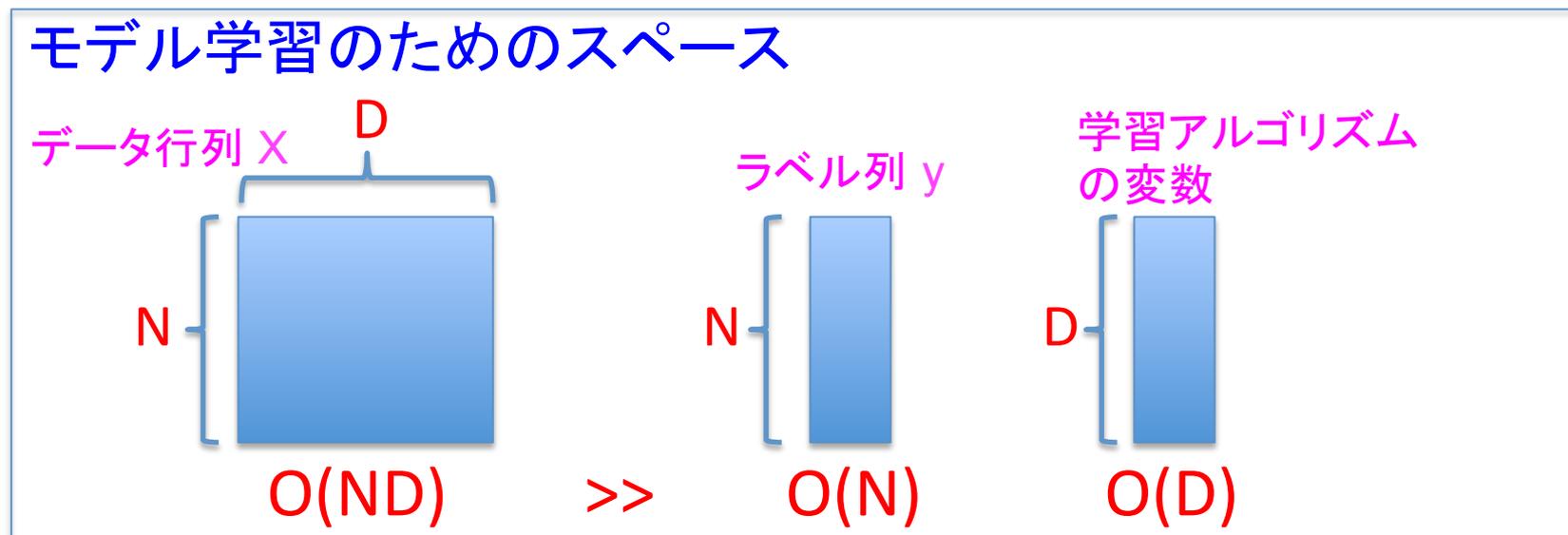
【学生優秀プレゼンテーション賞対象】

※サッカー試合データはデータスタジアム株式会社が提供

解釈可能な統計モデルをスケーラブルに学習させるための データ圧縮法 **Poster id : D1-44**

田部井靖生(東工大・JST-PRESTO) Joint work with
西郷浩人, 山西芳裕 (九大), Simon, J. Puglisi (ヘルシンキ大)

KDD'16のResearch Trackに採択

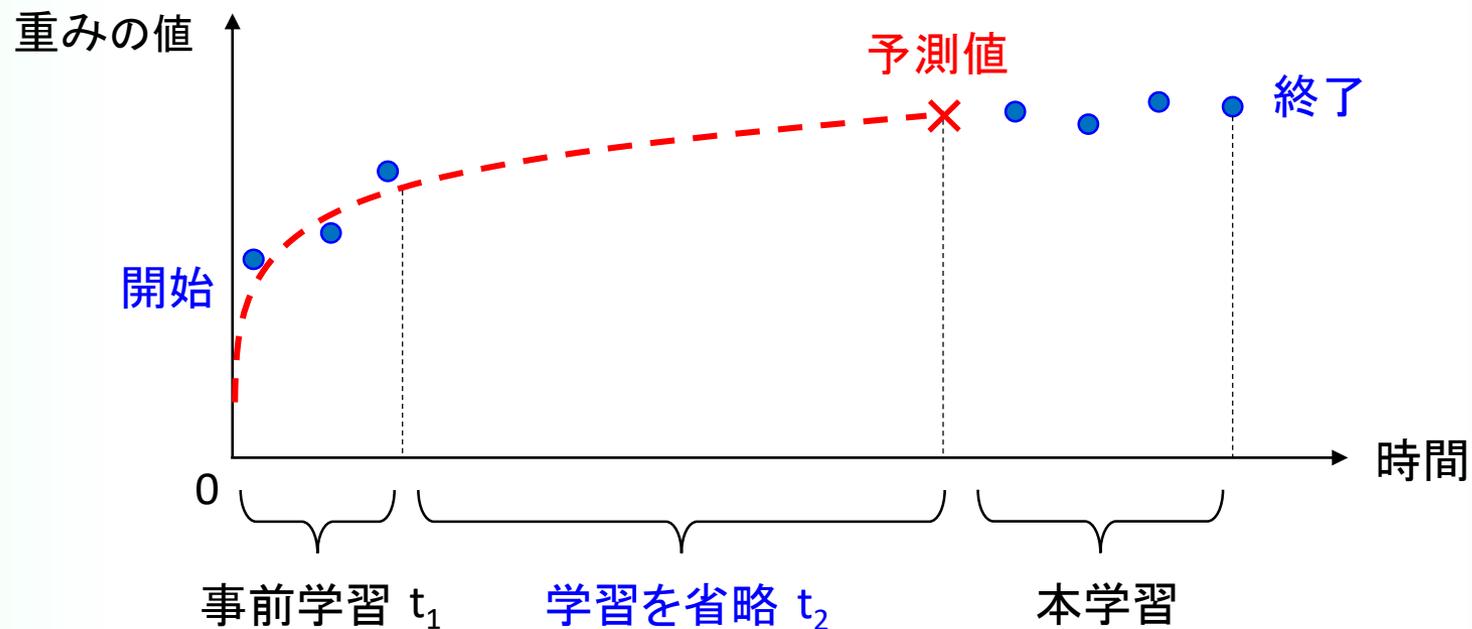


- データ行列のメモリに比べてラベル列と学習アルゴリズムのメモリは圧倒的に小さい
- 圧縮された行列上で学習アルゴリズムを動作できれば、スケーラブルにモデル学習可能
- 文法圧縮されたデータ行列上でPLS回帰モデル学習アルゴリズムを実装 (例：125GBのデータを**4GB**に圧縮可能)

D1-46 パラメータの漸近予測による ニューラルネットワーク学習の高速化

学生優秀プレゼンテーション賞対象

高瀬朝海・栗原正仁・小山聡（北海道大学 情報科学研究科）



各重みの時系列データを曲線で近似し、予測値を計算



学習を部分的に省略することで、学習時間を短縮

[D1-47] Riemannian stochastic variance reduced gradient

Hiroyuki Kasai (The University of Electro-Communications)

Hiroyuki Sato (Tokyo University of Science)

Bamdev Mishra (Amazon Development Centre India)

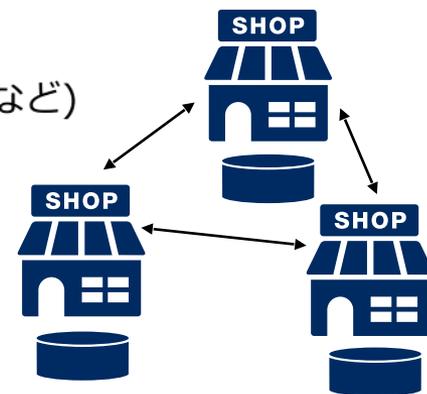
- ▶ Propose **Riemannian SVRG (R-SVRG)**, and
 - ▶ Extend **SVRG** in the **Euclidean** into **Riemannian manifolds**.
- ▶ Give two analyses;
 - ▶ **Global convergence** analysis, and
 - ▶ **Local convergence rate** analysis.
- ▶ For two settings;
 - ▶ **Exponential mapping** and **parallel translation**, and
 - ▶ **Retraction** and **vector transport**.

谷本 啓, 本橋 洋介 (NEC)

- 目的：タスク数が非常に多い場合の効果的学習
- 既存手法：各タスク間の関係を同時に学習 ([Argyriou 14] など)

$$\hat{W} = \arg \min_{W, L} \text{Loss}(W) + \lambda_1 \text{tr}(W L W^T) + \lambda_2 \text{tr}(L^{-1})$$

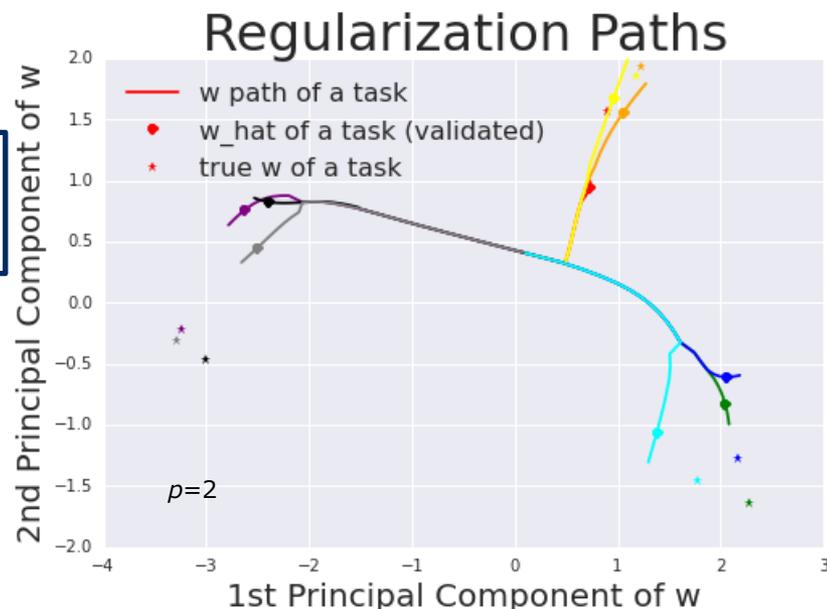
⇒ L の最適化は大変！ (各ステップ $O(\text{タスク数}^3)$ の計算量)



- 提案手法：Convex Fusion Penalty

$$\hat{W} = \arg \min_W \text{Loss}(W) + \lambda \sum_{i \neq j} \|w_i - w_j\|_p$$

- ✓ 事前知識不要 (導入も可能)
- ✓ 任意のクラスタ数に集約可能
- ✓ $p \geq 1$ で $O(T^2)$ 、 $p=1$ で $O(T \log T)$



D1-50:ラベルなしデータを活用した最新の分類器学習法

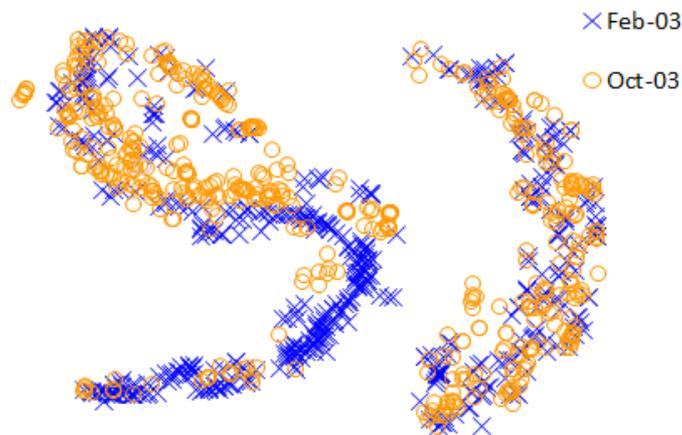
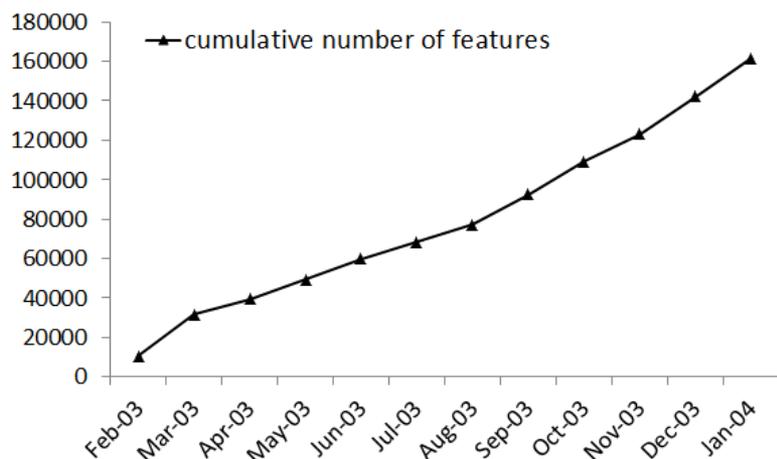
熊谷充敏 (NTT), 岩田具治 (NTT)



NTT Confidential

背景:

- 多くのドメインで分類器の性能は経時劣化してしまう
- 主要な要因は、**新たな特徴量の増加**と**データ分布の変化**

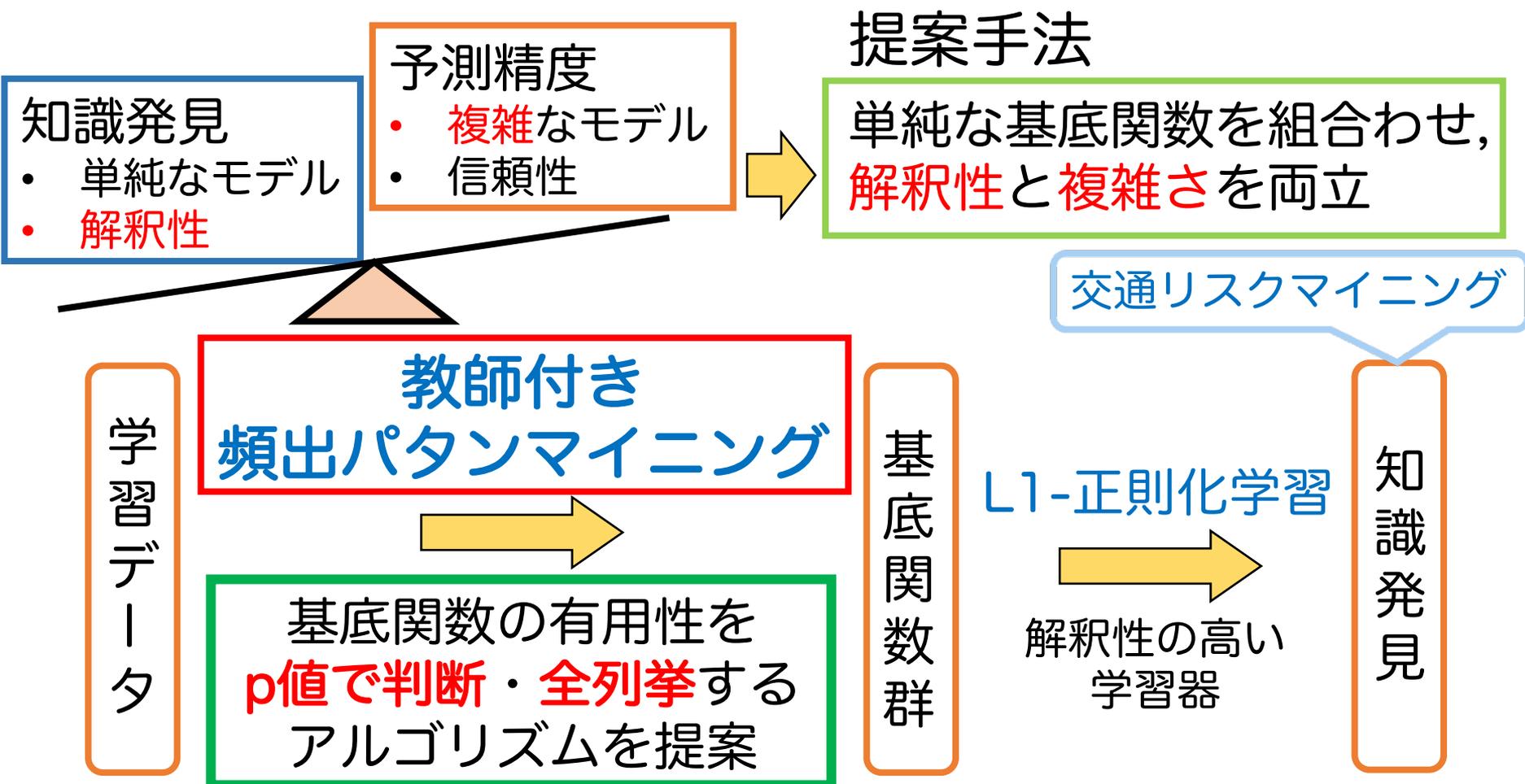


事前に得られたラベルありデータに加え、最新時刻に得られた**ラベルなしデータ**を用いて、**両問題を解決する分類器学習法**を提案する

キーワード: 欠損値補完、密度比推定、半教師あり学習、転移学習

頻出パターンマイニングに基づく 線形予測基底の選択

D1-51 李太斗, 松島慎, 山西健司 (東大, CREST)



D1-52 機械学習を用いた表情判定のための Action Unit最適化

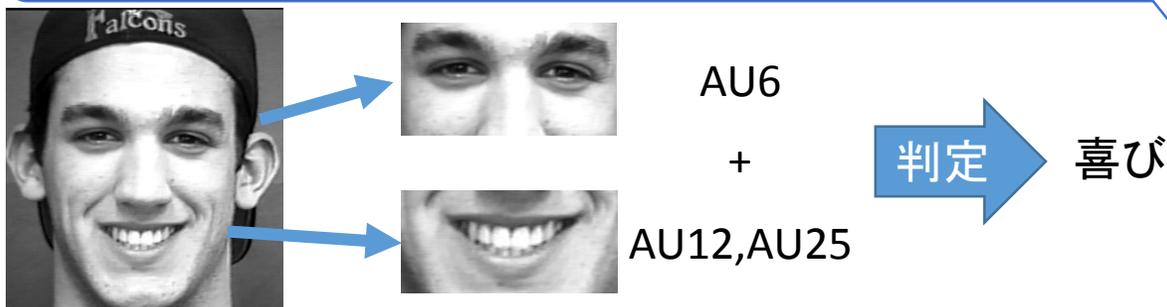
大澤 愛喜, 宇都 雅輝, 植野 真臣 (電気通信大学)

顔表情認識のアプローチ

1. 顔の特徴量 (Action Unit: AU) を抽出



2. 推定されたAUに基づき表情を判定



AUから表情の推定

既存手法

恣意的に選択したAUに基づき推定

➡ 推定精度が最適化される保証はない

本研究の目的

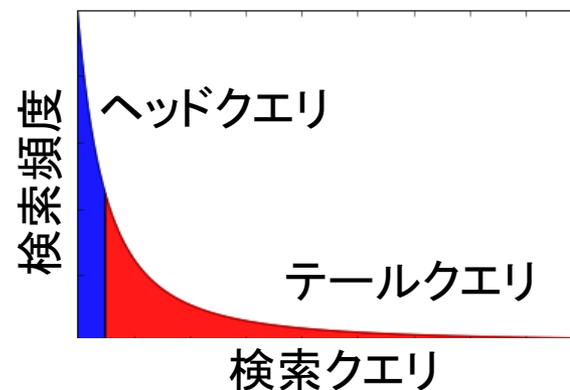
機械学習手法を用いて、
予測を最適化するAUを選択

検索キーワードの検索頻度を考慮した 商品検索のランキングモデルの構築

齋藤 祐樹 (ヤフー株式会社)

問題設定

- 商品検索におけるランキング問題
- 検索キーワードの検索頻度は非常に **ロングテール** な分布を持つ
- 検索数の多い **ヘッドクエリ** の精度向上が全体の精度改善につながる



提案手法

- 学習時に検索頻度に応じて勾配を **調整** する

勾配の更新式:
$$h_k(x) = \frac{kh_{k-1}(x) + \eta c(x) g_k(x)}{k+1}$$

- 実データを用いて手法の効果を確認

D1-54: 単語の分散表現の逐次的学習

鍛治伸裕 小林隼人(ヤフー株式会社)

- 既存の単語の分散表現学習は逐次学習が困難
 - ツイートなどの増加し続けるデータからの学習は非効率
- Skip-gram model with Negative Sampling (SGNS)をオンライン学習アルゴリズムに拡張
 - 単語の分散表現学習としては初のオンラインアルゴリズム

効率的な実装: weighted reservoir sampling による実時間処理

理論的な解析: オンライン最適解がバッチ最適解に確率収束することを証明

大規模な実験: 理論検証、バッチSGNSとの比較

Sparse-view X線CT



少ない投影データからCT画像の再構成を目指す

圧縮センシング

$$\hat{f} = \underset{\theta}{\operatorname{argmin}} \{ \|g - Cf\|_2^2 + \lambda_{tv} \|TV(X)\|_1 \}$$



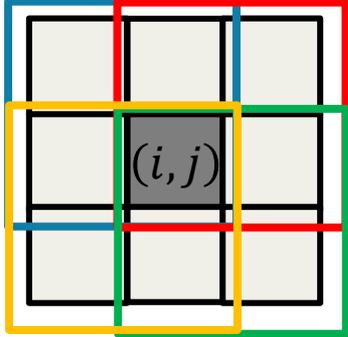
提案手法

複数のTV-priorを組み合わせる

1. anisotropic-TV

$$\|f\|_{anisotv} = \sum_{i,j} (|\nabla_x f|_{i,j} + |\nabla_y f|_{i,j})$$

2. Pseudo-TV



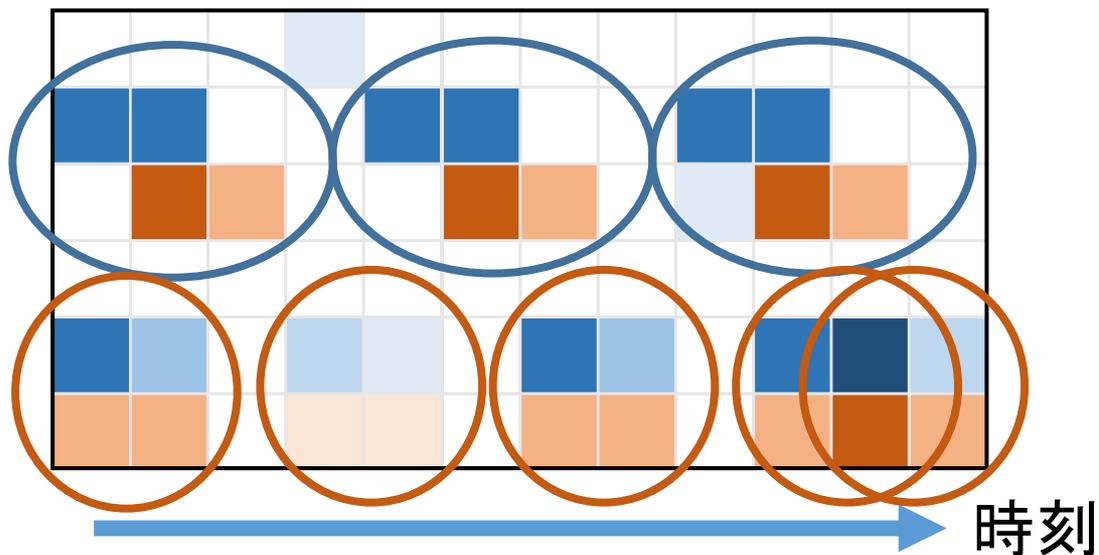
結果

画像によって、よく効くTV-priorが異なる

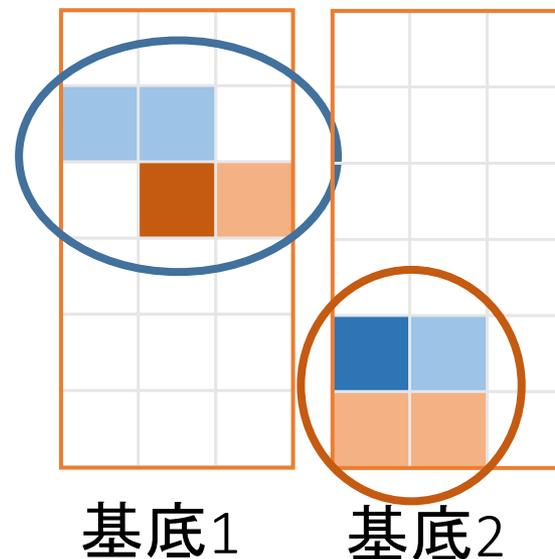
畳み込み半非負値行列因子分解と, **D1-56** 記述長最小原理に基づく基底数推定法

学生優秀プレゼンテーション賞対象 ○鈴木 惇¹, 山西 健司^{1,2} 1: 東京大学 2: CREST

Given: 多次元時系列データ



Get: 時間連続な特徴



- ・パラメータ推定アルゴリズムの提案
- ・基底数推定法の提案

工藤康統(慶應義塾大学), 楠本充(Preferred Networks(株)), 藤田康博(Preferred Networks(株))

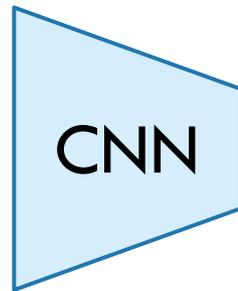
Convolutional Neural Networkを用いた強化学習
による3Dレーシングゲーム上での自動運転の獲得



事前に模倣学習を行うことが深層強化学習においても有効



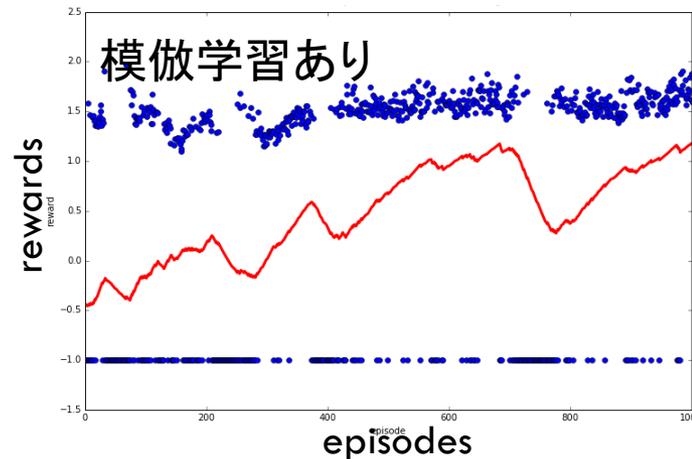
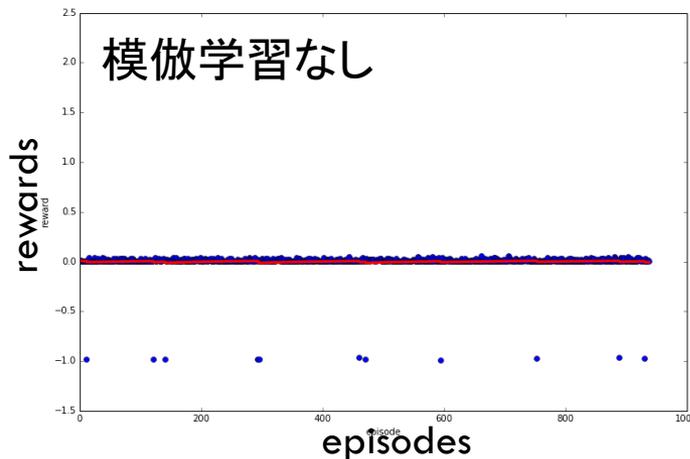
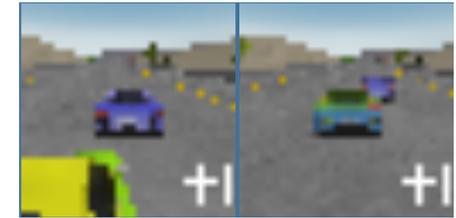
入力画像



制御

□ single-play

□ self-play



学生優秀プレゼン
テーション賞対象

D-58 劣モジュラ正則化雑音除去の自由度

南 賢太郎 (東大)

Our Goal

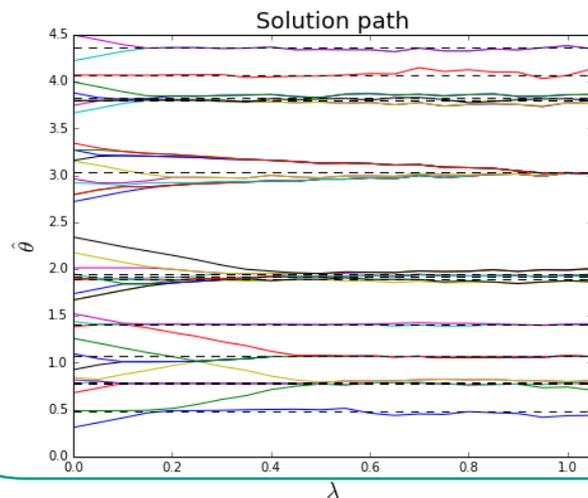
- Lovász拡張型正則化の理論解析
- 特に**自由度**に興味 (リスクの不偏推定)

$$\min_{\theta \in \mathbb{R}^n} \frac{1}{2} \|y - \theta\|_2^2 + \lambda \hat{f}(\theta)$$

Idea

分割ごとに定数である解が出やすい

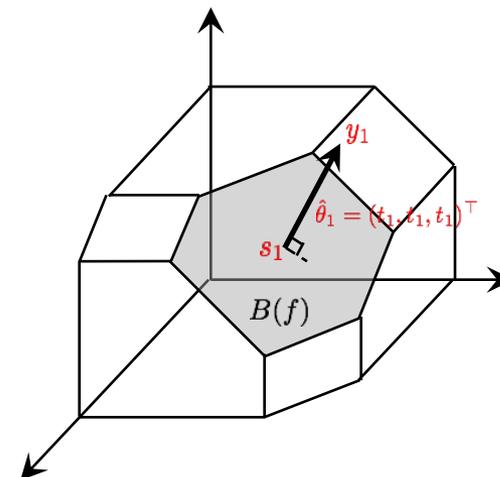
【疑問】 分割数 = モデルの実効次元？



Theory

YES! 一般に $\mathbb{E}[\text{分割数}] = \text{自由度}$

【証明】 基多面体の法錐に着目



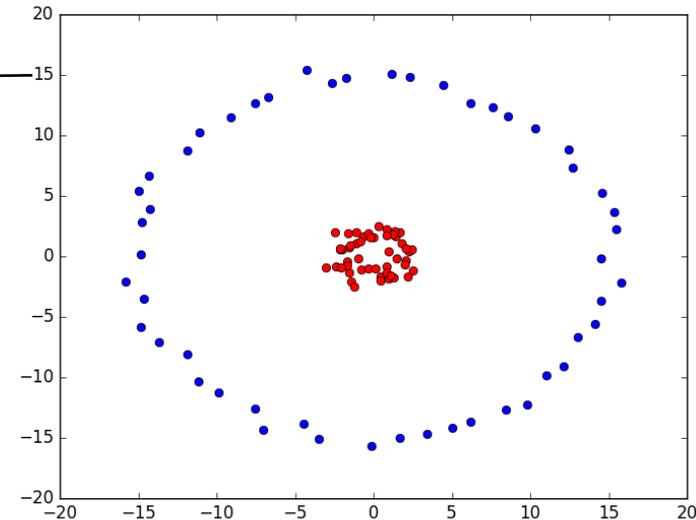
“A Gaussian Process Model for Non-Convex Probabilistic Clustering”

Motivation:

- Good clustering performance like spectral clustering, even for **non-convex shaped** clusters.
- In contrast to spectral clustering, enable Bayesian model selection of hyper-parameters.

Idea:

- Use a Multivariate Gaussian Processes Regression model similar to **Gaussian Process Latent Variable Model (GP-LVM)**.



Inference:

- **Relatively Fast.**
Reduced complexity of MCMC step: $O(n^3) \rightarrow O(n^2)$

Preliminary Experiments:

- **Look promising** … more see poster **D1-59**

Daniel Andrade, Graduate University of Advanced Studies (SOKENDAI)*,
Kenji Fukumizu, The Institute of Statistical Mathematics.

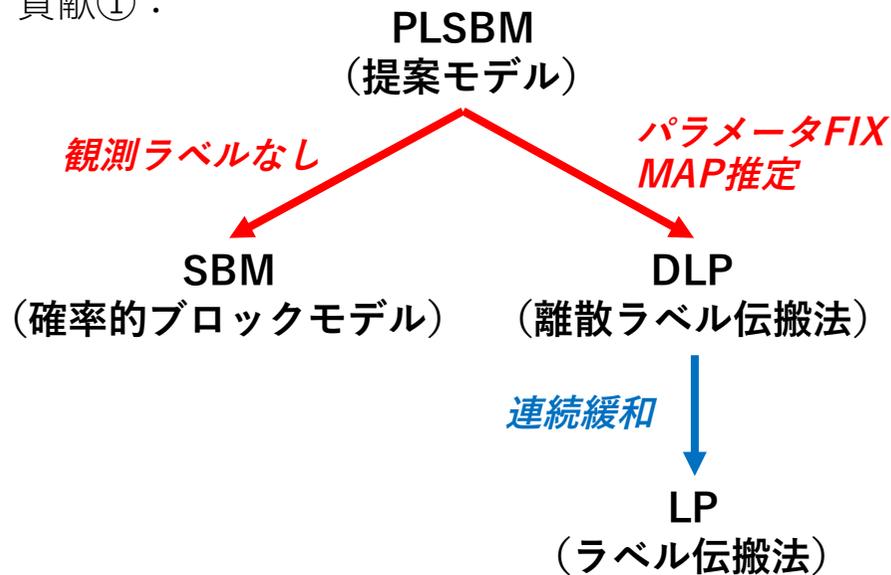
* The first author is also a member of NEC Data Science Research Laboratories

ラベル伝播法はいつ失敗するか？ネットワーク生成モデルとしての解釈

D1-60 山口 祐人 (産総研)
林 浩平 (産総研)

貢献①：ラベル伝搬法と確率的ブロックモデルとの理論的関係性を示す
貢献②：ラベル伝搬法が失敗する4つの条件を理論的・実験的に示す

貢献①：



貢献②：

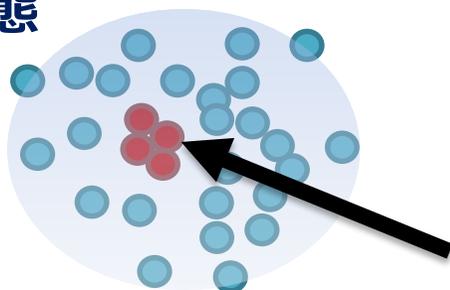
- 条件1：異なるラベル同士が接続
- 条件2：クラスタごとの接続密度が異なる
- 条件3：ラベルごとの数が異なる
- 条件4：観測ラベルが誤っている

D1-61: 教師なし異常検知器に対する誤検知情報のフィードバック手法

山中 友貴, 山田 真徳 (NTTセキュアプラットフォーム研究所)

■ VAE等の再構成誤差に基づいた教師なし異常検知

正常状態

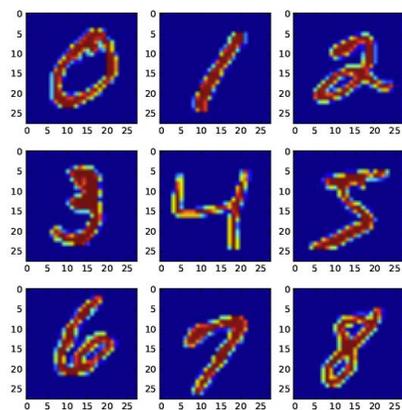


正常状態に
まぎれた異常

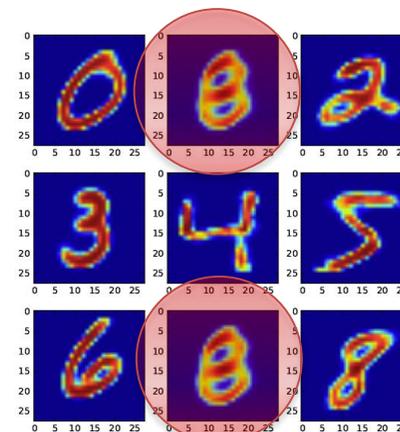
検知漏れ！

正常状態にまぎれた異常状態の
誤検知を防止したい

■ 少数の異常データを用いて 異常状態の学習を阻害する方法を提案



損失関数を改変し
学習を阻害

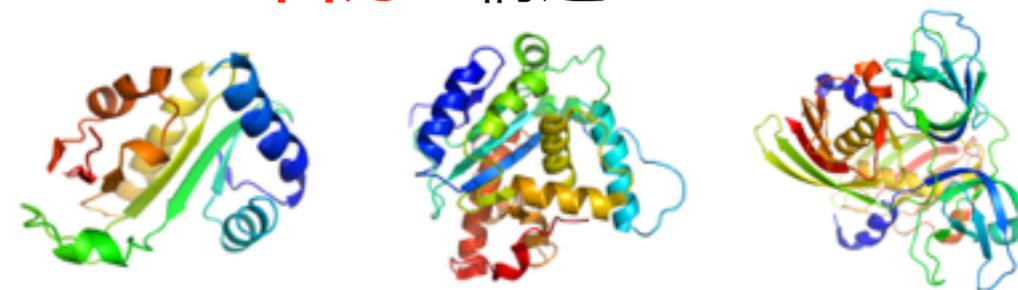


進化情報と物理情報を用いたタンパク質の立体構造予測と立体構造情報からのタンパク質生成

入力：タンパク質（任意長のアミノ酸配列）

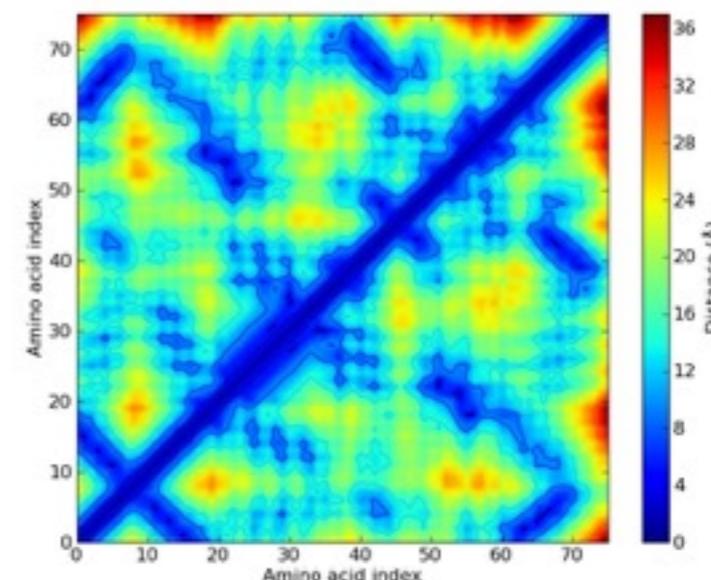
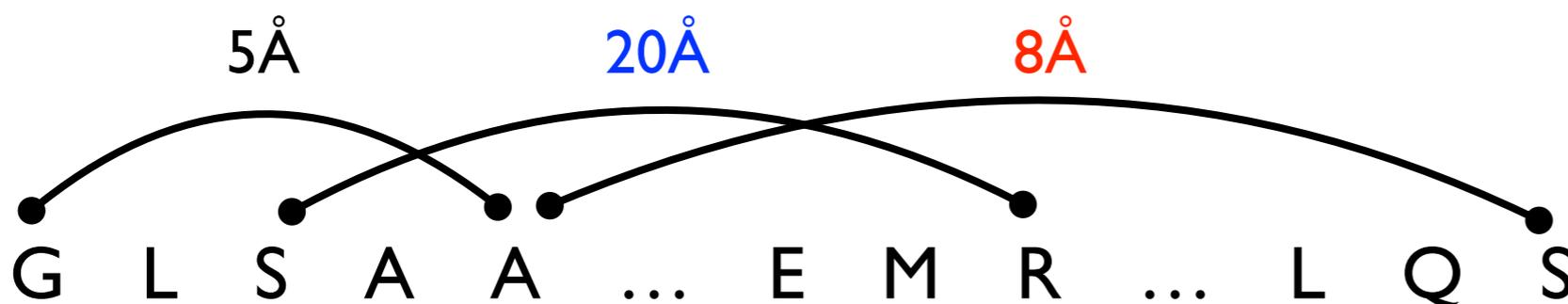
出力：構造ラベル

GLSAAQRQVIA · · · ISGLQS



入力：タンパク質中の全アミノ酸ペア

出力：全アミノ酸間の距離



入力：3次元内部座標系列（Z-matrix）

出力：アミノ酸配列

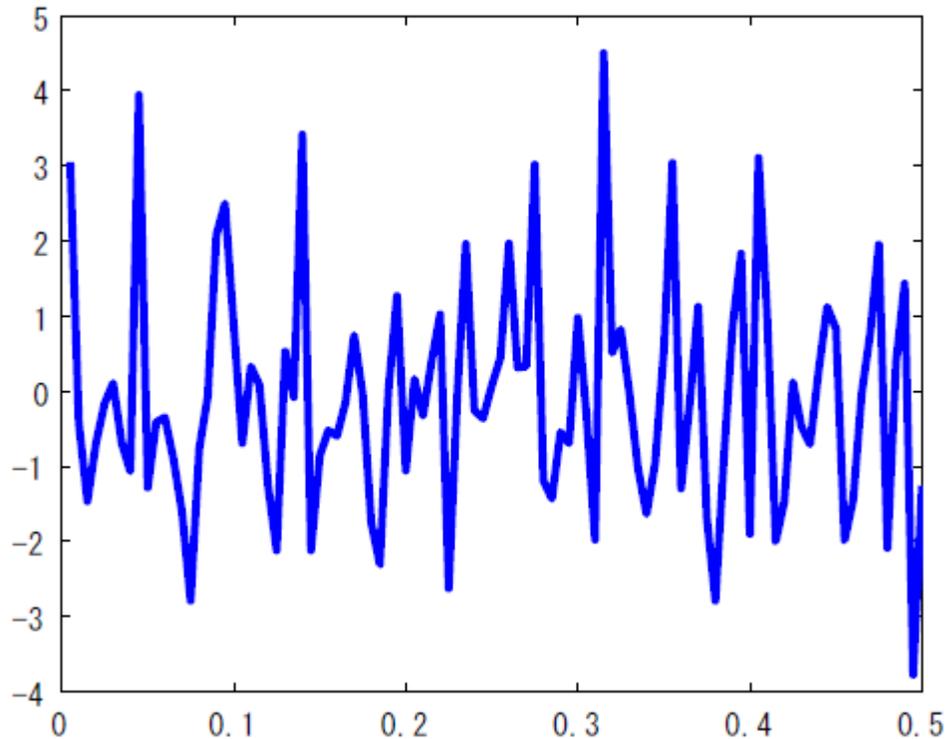


GLSAAQRQVIA · · · ISGLQS

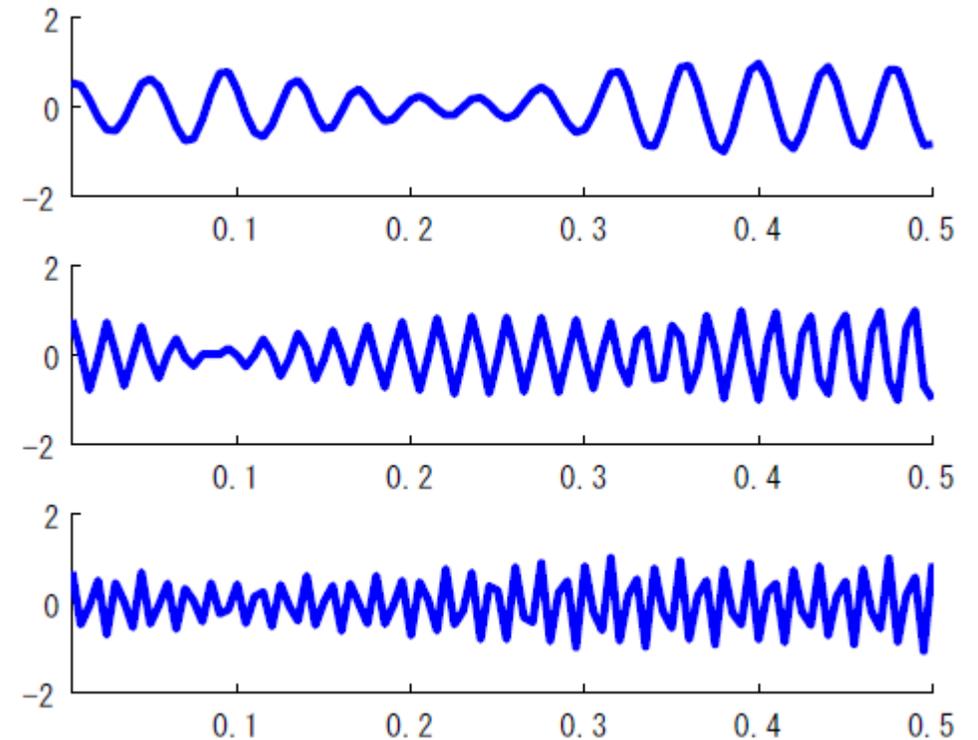
タンパク質の**進化情報**、螺旋構造とシート構造が形成する**物理的制約**、アミノ酸の間接的な**相互作用**を考慮しモデル化したニューラルネットワークを提案

D1-63: 時系列データの周期成分分解と位相推定

松田 孟留(東京大・情報理工), 駒木 文保(東京大・情報理工&理研・脳センター)



提案手法



時系列データの背後にある振動子を取り出す手法を開発
神経電位データ・黒点データなどに適用

学生優秀プレゼンテーション賞対象

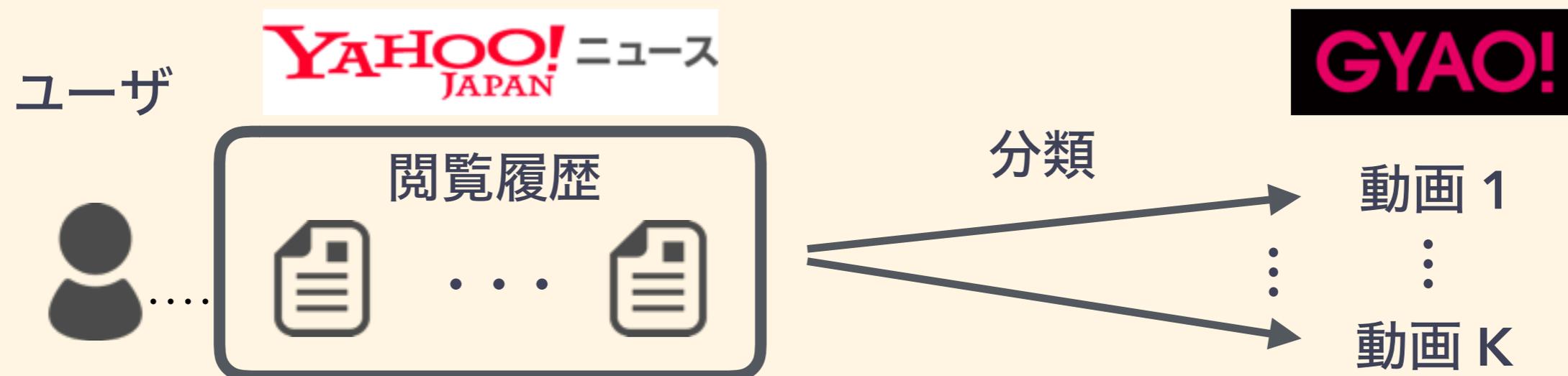
ニュース・動画サービス間の クロスドメイン推薦における課題

D1-64

金川 平志郎¹, 小林 隼人², 清水 伸幸², 田頭 幸浩², 鈴木 大慈¹

¹東京工業大学 ²ヤフー株式会社

(学生優秀プレゼンテーション賞対象)



ニュースは使うが、動画を見たことのないユーザへ推薦

アプローチ：ニュース履歴の動画への分類

- ドメイン適応 + NN (Domain-Adversarial NN, ICML2015)
- 共通のユーザが不要

D1-65: Stochastic Primal Dual Coordinate Method における適応的サンプリング

柴垣 篤志, 竹内 一郎 (名工大)

目的: SPDC, Doubly SPDC の高速化

手法:

一様ではなく適応的に確率分布を変更

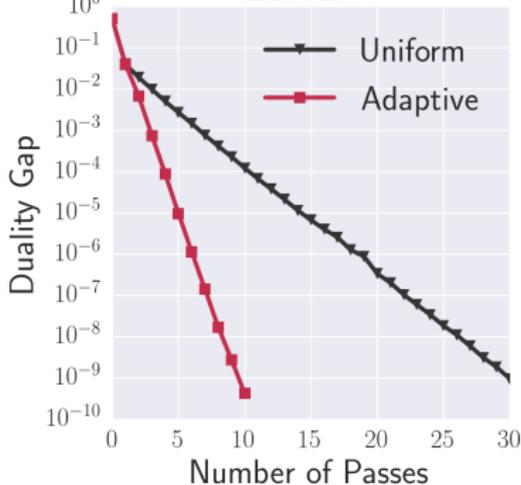
e.g.,

- ▶ 各座標の最適性条件の違反度に応じた分布

学生優秀プレゼンテーション賞対象

Elastic net + Smoothed hinge loss

real-sim ($n=72201$, $d=20958$) $\lambda = 1e-3$, $\lambda_{\max} = 1.45 * 10^{-4}$
minibatch size = 4

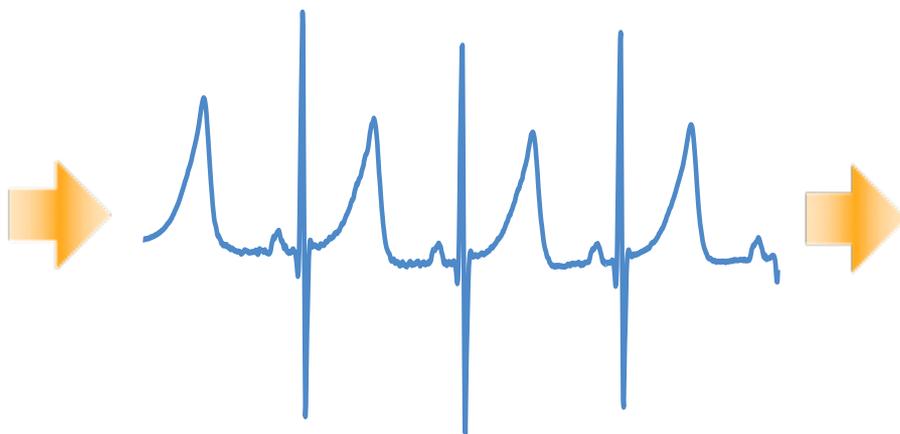


心拍変動解析と統計的プロセス管理に基づいた てんかん発作予知アルゴリズム

○藤原幸一, 加納学 (京都大学), 宮島美穂(東京医科歯科大学),
山川俊貴(熊本大学)



心拍センサ



心拍データ



アプリでの解析

D1-66にて実機を用いたデモを行います

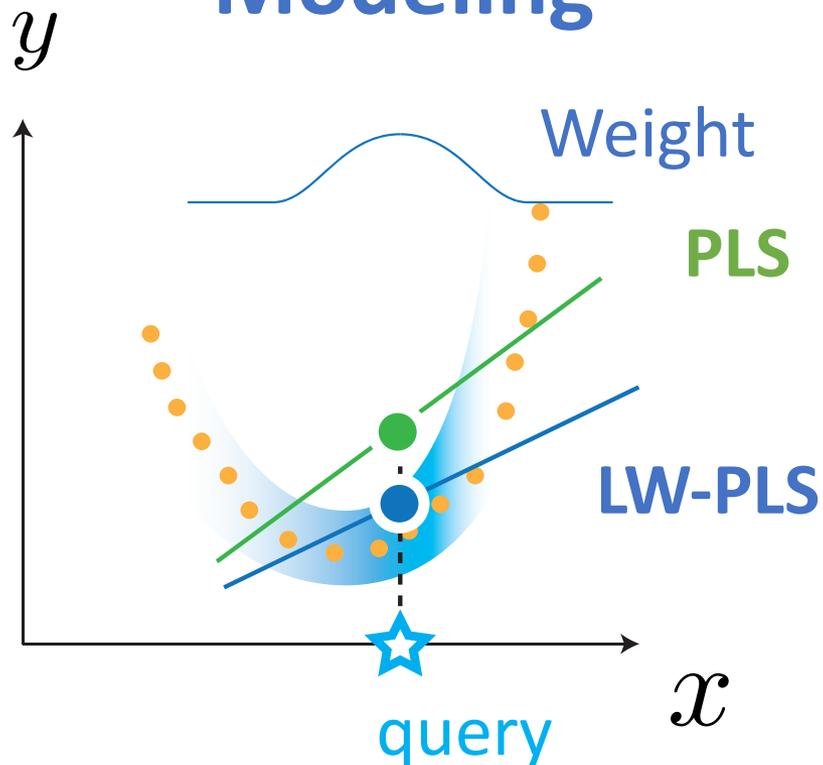
D1-67 局所PLSを用いた欠損RRI補間

○ 鎌田 啓輔*, 藤原 幸一*, 山川 俊貴**, 加納 学*

学生優秀プレゼンテーション賞候補

* 京都大学, ** 熊本大学

Just-In-Time Modeling



Data missing

RRI : 心拍データ

