

学習数理情報学研究室の紹介

東京大学 大学院情報理工学系研究科
数理情報学専攻
第6研究室

<http://www.ibis.t.u-tokyo.ac.jp/>

2018/10 /26

学習数理情報学研究室

- 機械知能の本質を数理で捉える -

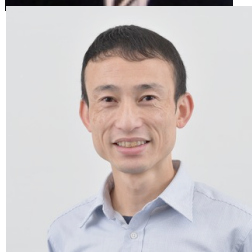


教授 山西健司

専門: 情報論的学習理論、データマイニング、異常検知

講義: 機械学習の数理(夏:学部)

情報論的学習理論(冬:大学院)



准教授 大西立顕 (SICT本務、数理兼務)

専門: 経済物理学、社会物理学、ビッグデータ解析

講義: グローバル・クリエイティブリーダー講義III

(情報システム論)(夏:大学院)



准教授 鈴木大慈

専門: 機械学習, 数理統計学, 確率的最適化

講義: 確率数理工学(夏:学部),

確率数理要論(冬:大学院)



助教

二反田篤史

専門: 確率的最適化

深層学習



特任助教

木脇太一

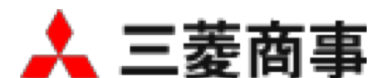
専門: ニューラルネット

講義: データマイニング

実践演習

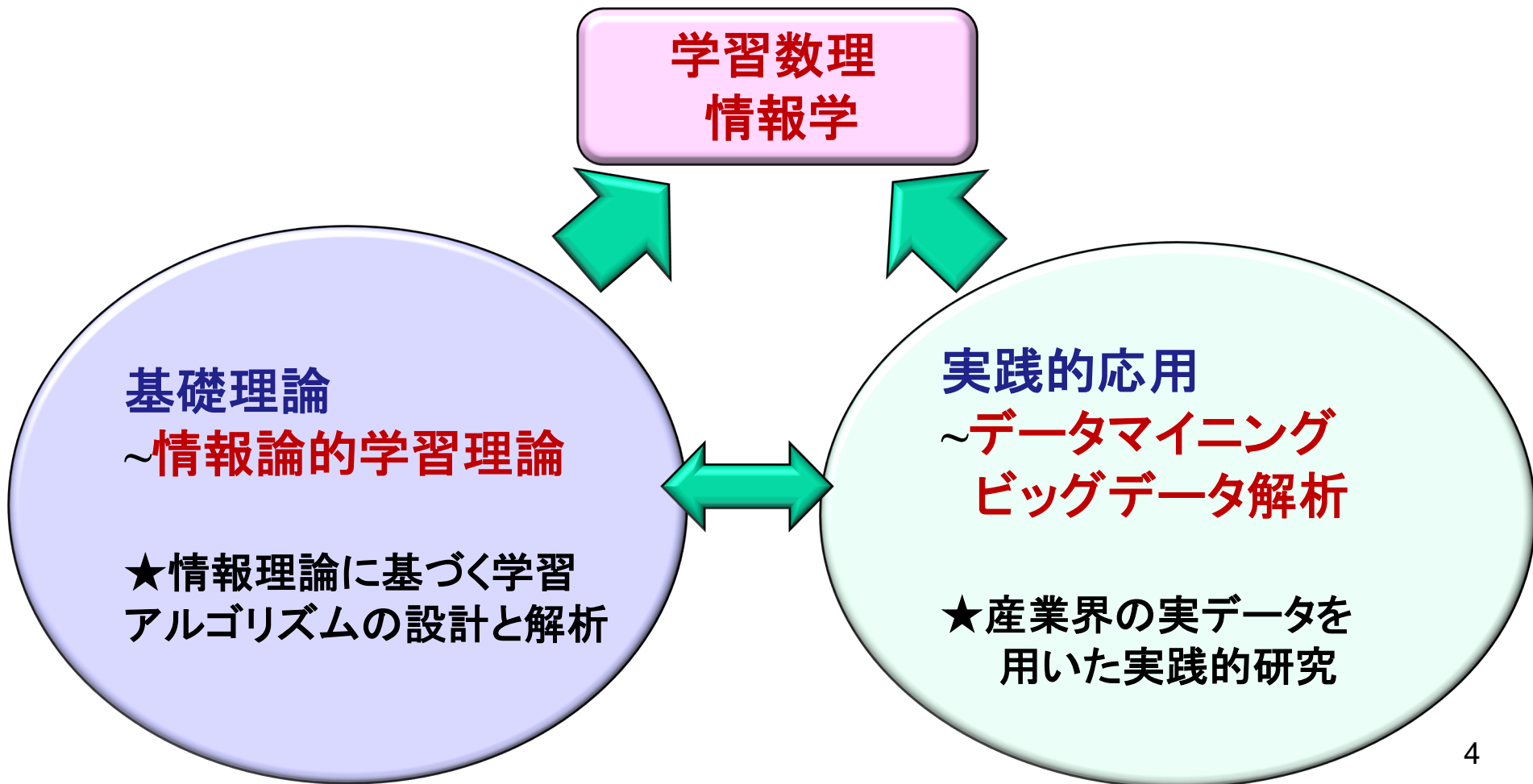
研究室の特徴

- しっかりとした基礎理論に立脚しつつ、現場に生きる数理情報学をめざします
- JST CREST ビッグデータ基盤「複雑データからのディープナレッジの発見と価値化」の研究拠点です
- 幅広い産業界との連携のもとで、研究を進めます
以下の企業とはすでに協力を開始しています。



学習数理情報学における両輪

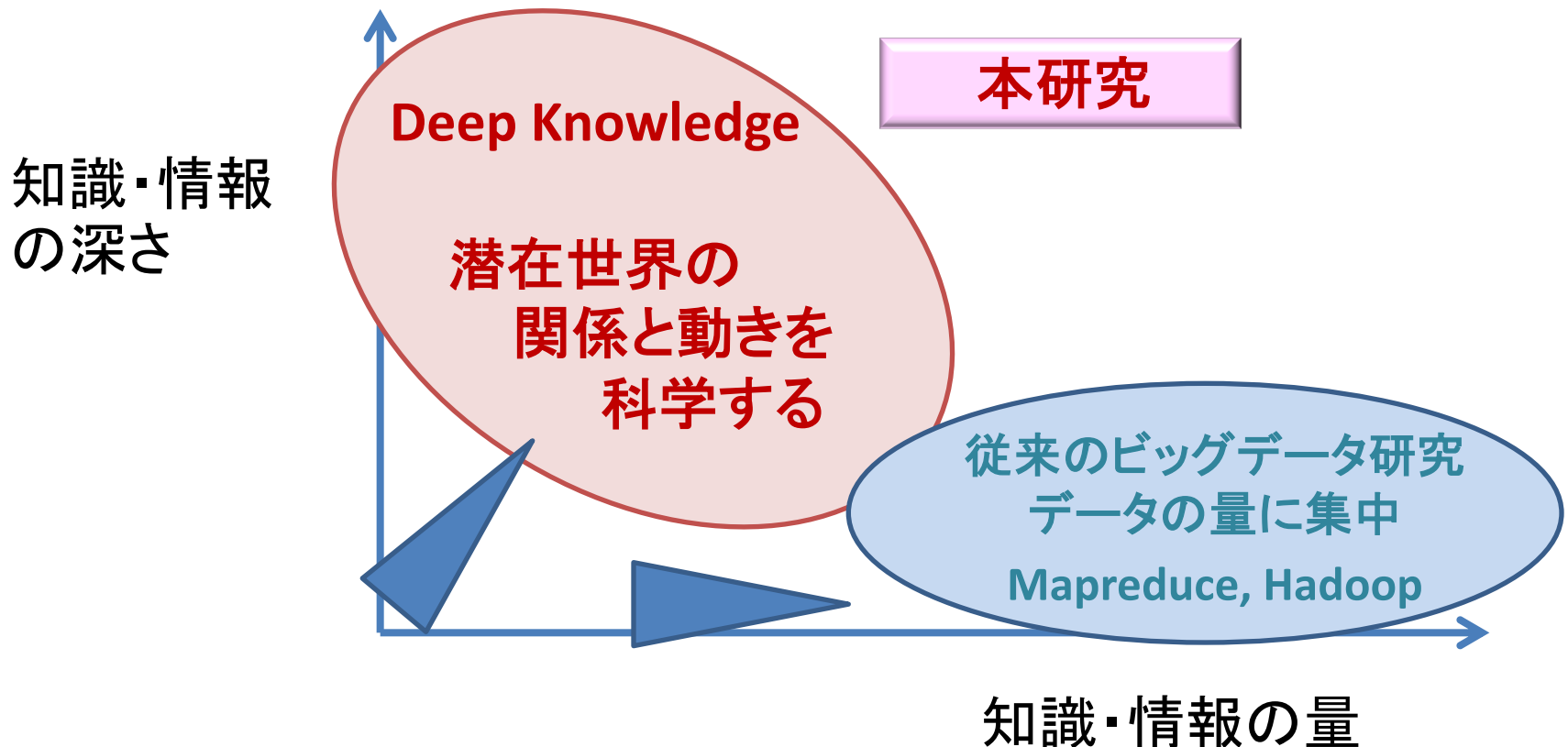
情報論的学習理論とデータマイニング実践を両輪とする



日本学術振興機構 CREST

「複雑データからのディープナレッジの発見」の研究拠点(2013～2019)

知識の深さに重点を置き、データに潜在するDeep Knowledgeを発見し、valueを引き出す新しい方法論を開発し、体系化する



山西研究室の概要

コア技術: **情報論的学習理論**
~ **MDL(minimum description length) 原理**

目標: **ディープナレッジの発見**

潜在的
ダイナミクス
Velocity

基礎理論

潜在構造
最適化
Variety

主要応用

緑内障
進行予測

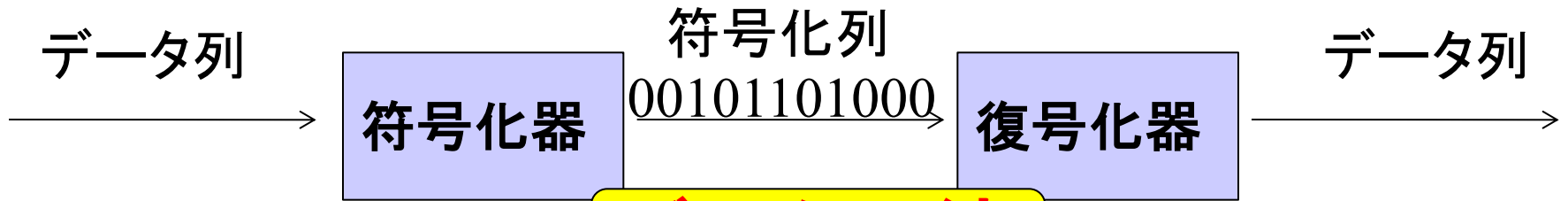
交通リスク
マイニング

マーケッ
ト解析

SNS
解析

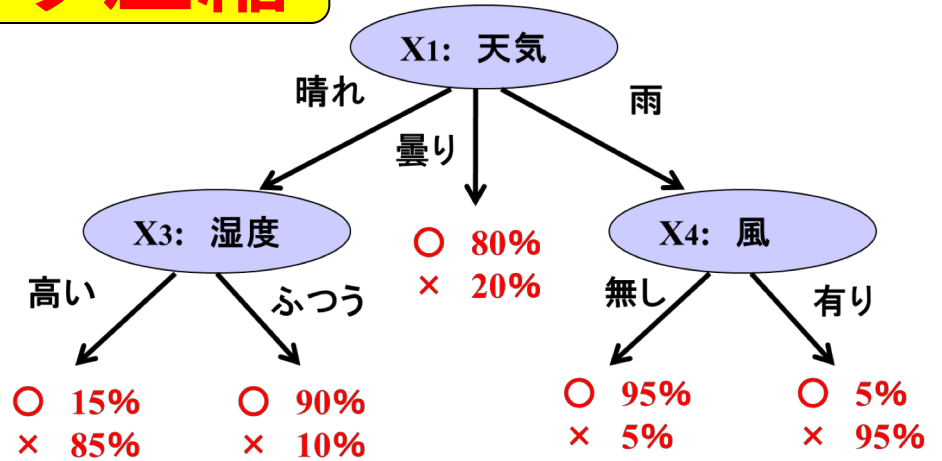
情報論的学習理論とは

学習とは最もデータ圧縮できるモデルを見つけることである



データ圧縮

X1	天気	X2	温度	X3	湿度	X4	風	Y	ブレ
	晴れ		暑い		高い		無し		×
	晴れ		暑い		高い		有り		×
	曇り		暑い		高い		無し		○
	雨		暖かい		高い		無し		○
	雨		涼しい		ふつう		無し		○
	雨		涼しい		ふつう		有り		×
	曇り		涼しい		ふつう		有り		○
	晴れ		暖かい		高い		無し		×
	晴れ		涼しい		ふつう		無し		○
	雨		暖かい		ふつう		無し		○
	晴れ		暖かい		ふつう		有り		○
	曇り		暖かい		高い		有り		○
	曇り		暑い		ふつう		無し		○
	雨		暖かい		高い		有り		×
.....



$$\mathcal{L}(D^n | M) + \mathcal{L}(M) \Rightarrow \text{最小化 w.r.t. } M$$

[データ符号長] [モデル符号長]

確率的コンプレキシティ

MDL (Minimum Description Length) 原理

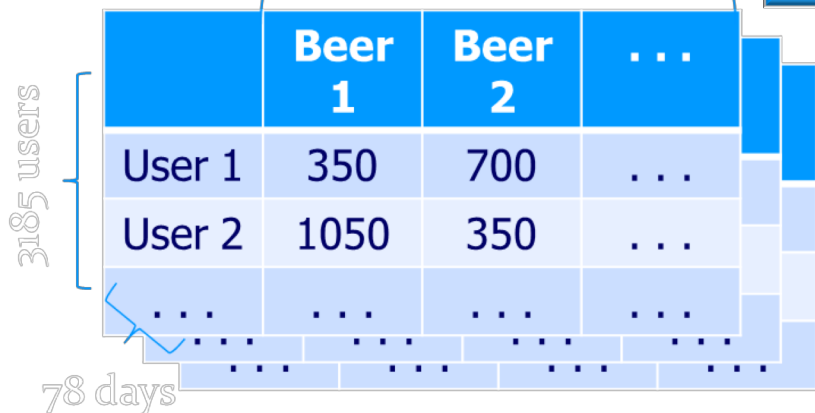
潜在的ダイナミクス：動的モデル選択

モデル（クラスター数）の時間変化をMDL原理で検知

[Yamanishi Maruyama IEEE IT2007,
Hirai Yamanishi KDD2012,, IEEE IT2013]

多変数時系列

14 kinds of beer



購買層クラスター系列

発泡酒派

変わらぬ派

プレミアム派

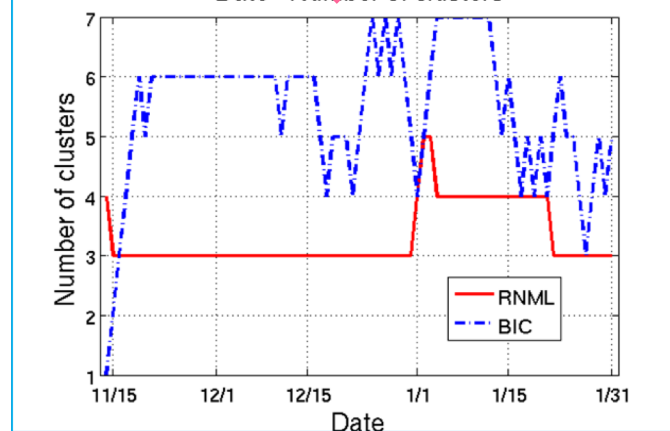
ベルギー
ビール派

Change

Change

Time

Date - Number of clusters



Inc.

最もデータ圧縮するモデル系列を求める
(MDL原理)

$$\sum_{t=1}^n (-\log p(x_t | x^{t-1}; \hat{\theta}_{t-1}, k_t)) + \sum_{t=1}^n \ell(k_t | k_{t-1})$$

$$\implies \min \text{ w.r.t. } (k_1, \dots, k_n)$$

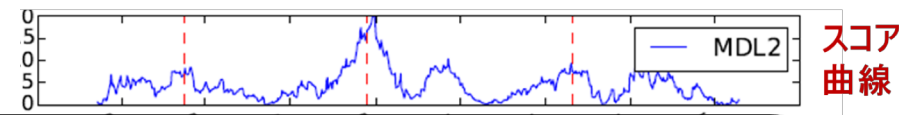
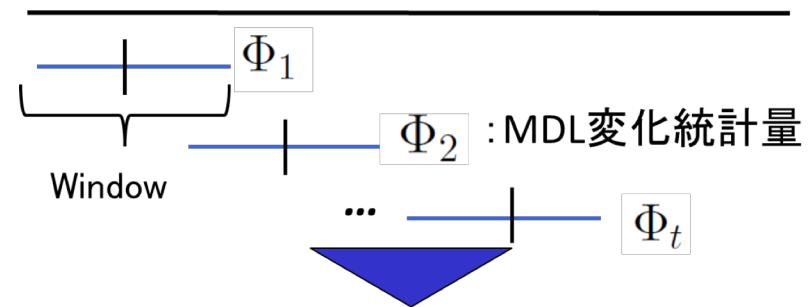
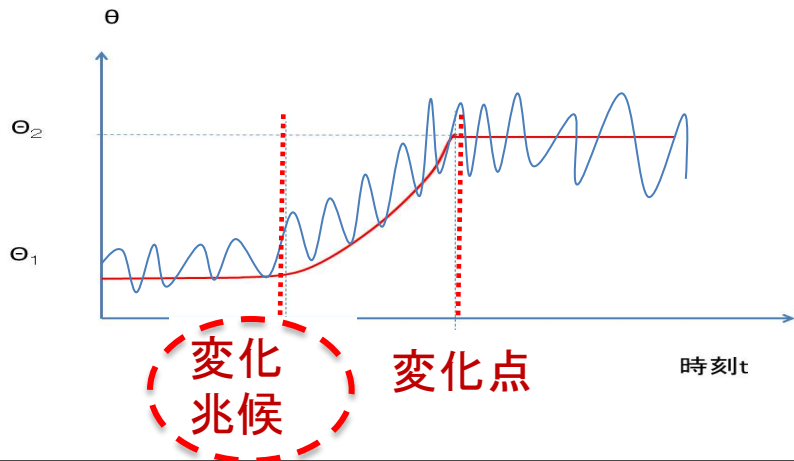
潜在的ダイナミクス: MDL変化統計量

データ圧縮度合いで漸進的変化の開始点(予兆)を検知

[Yamanishi, Miyaguchi BigData2016] [Kaneko, Miyaguchi, Yamanishi BigData2017]

[Yamanishi, Fukushima IEEE Trans Inform Theory 2018]

$$x^n = x_1 \dots x_n,$$



技術コア: MDL変化統計量によるオンライン変化検知を提案

MDL: Minimum Description Length

$$\Phi_t = \min_M L_{\text{NML}}(x_1^n; M) - \min_{M_1, M_2} \{L_{\text{NML}}(x_1^t; M_1) + L_{\text{NML}}(x_{t+1}^n; M_2)\}$$

変化した場合と
しない場合の
記述長の差

- ⇒
- ・固定窓で移動計算、漸進的変化検知 [BigData 2016]
 - ・可変窓で移動計算、検知精度を大幅改善 [BigData 2017]
 - ・工場のボイラー事故の予兆を検知(東レとの実証実験)

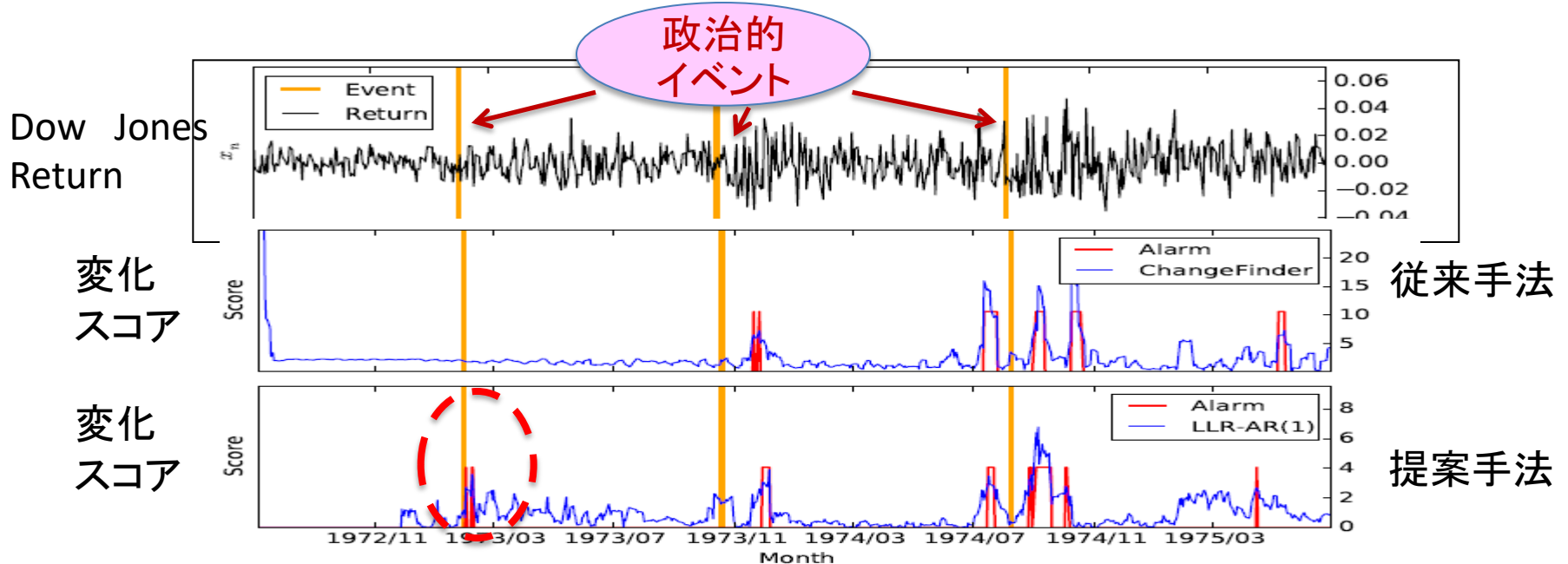
潜在的ダイナミクス：連続変化検知

連続変化をパラメタライズ、変化の兆候を情報論的な測度で検知

[Miyaguchi, Yamanishi DSAA2015, JDSA2016]

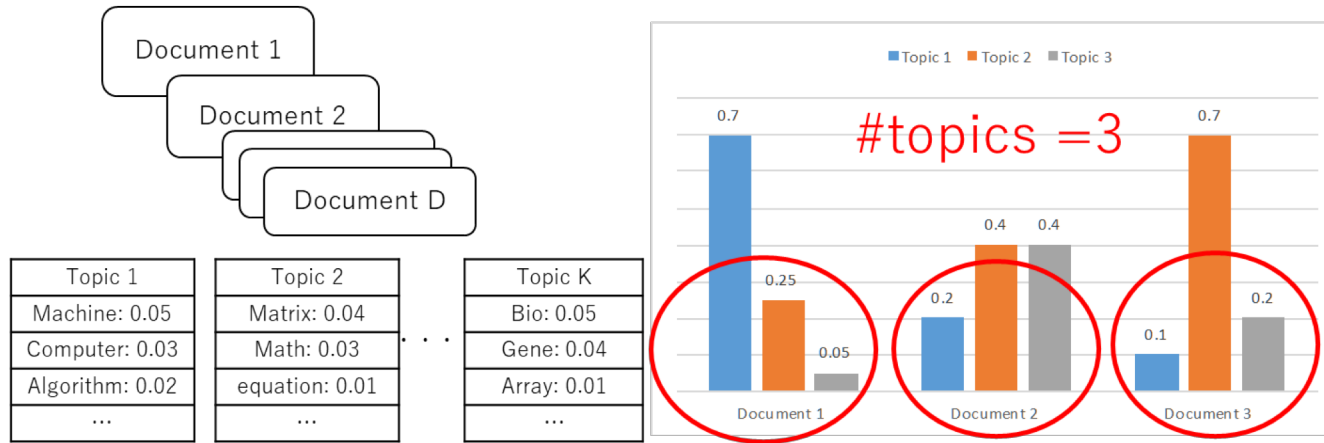
$$(\hat{\theta}_t, \hat{\delta}_t) = \arg \max_{\theta, \delta} \sum_{i=t-h}^{t+h} \log p(x_i | x^{i-1}; \theta + (i-t)\delta)$$

連続変化測度 $y_t \stackrel{\text{def}}{=} \sqrt{\hat{\delta}_t^T I(\hat{\theta}_t) \hat{\delta}_t}$ $I(\theta)$: Fisher情報量



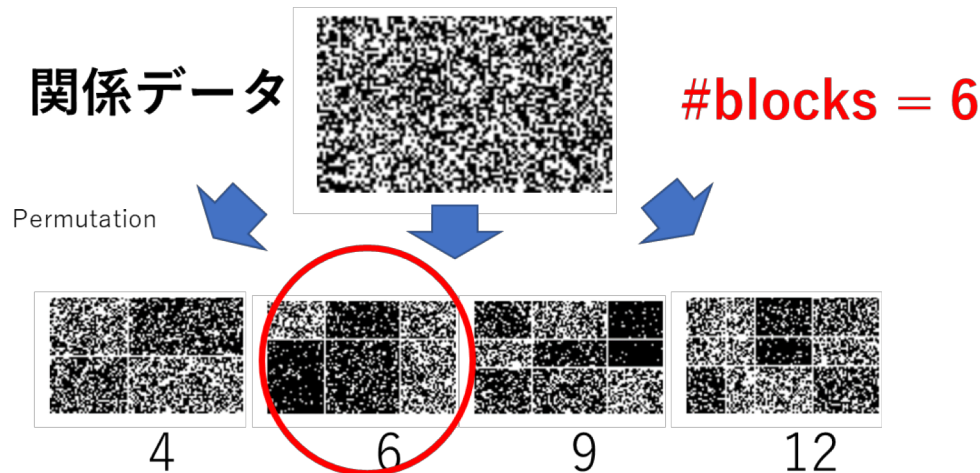
潜在構造最適化：潜在変数モデル選択

ドキュメントに含まれるトピック数Kは？



トピック
モデル

確率ブロックモデルのブロック数推定

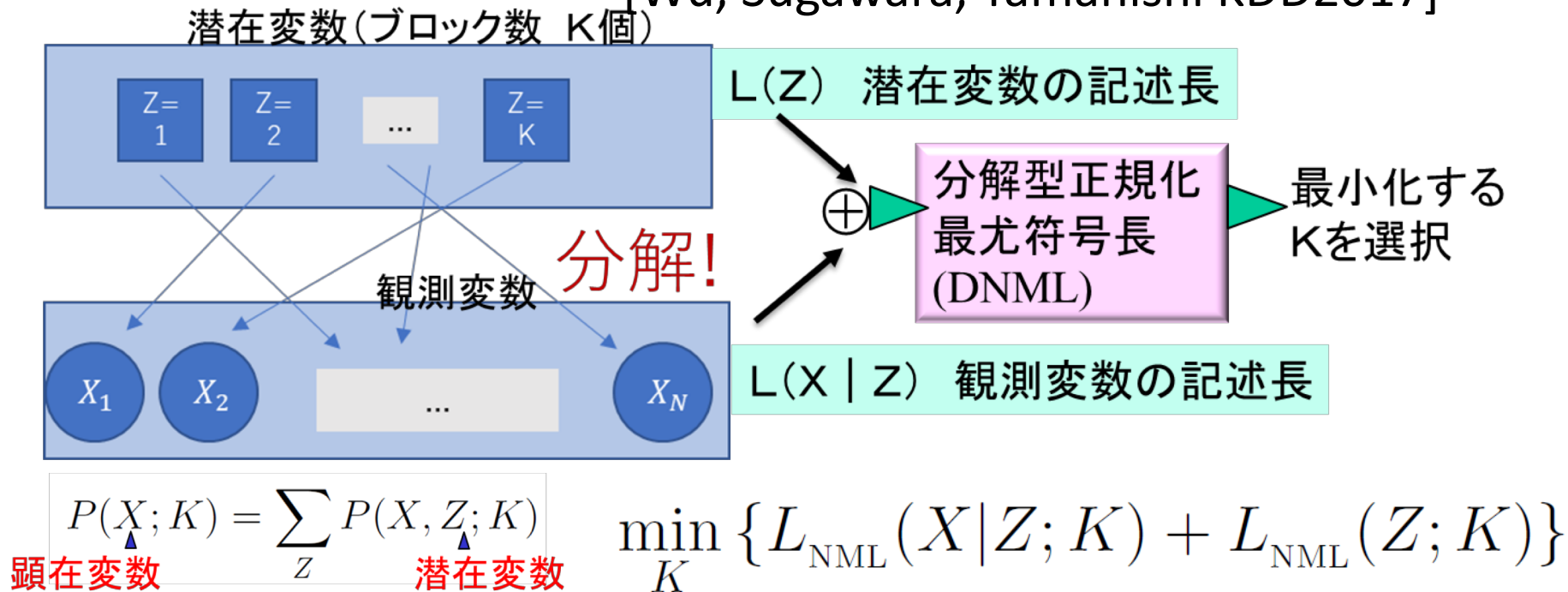


確率ブロック
モデル

潜在構造最適化：分解型NML規準 (DNML)

分解型正規化最尤符号長 (Decomposed Normalized Maximum Likelihood Codelength : DNML)

[Wu, Sugawara, Yamanishi KDD2017]



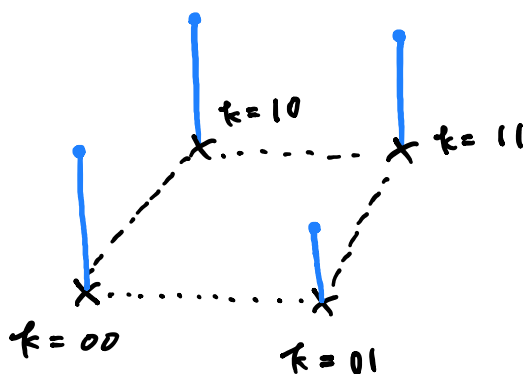
- 1) 効率計算可能
- 2) 一般の階層モデルに適用可能
- 3) 従来に比べて高精度かつロバストなモデル選択

潜在構造最適化：緩和的確率のコンプレキシティ

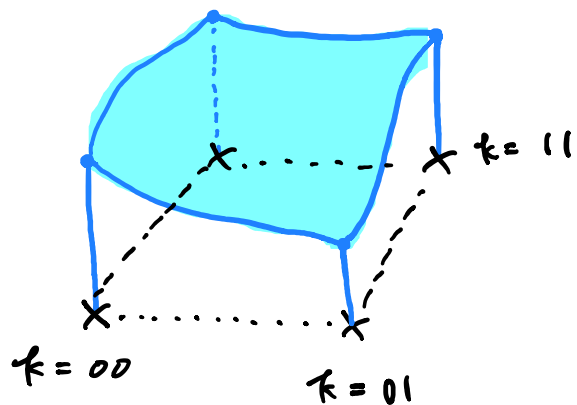
MDL規準を連続緩和してスパースモデルを効率的に選択

[Miyaguchi, Matshushima, Yamanishi SDM2017]

SC (Stochastic Complexity)



RSC (Relaxed Stochastic Complexity)



$$\text{RSC}(X; \lambda) = \underbrace{\min_{\theta} \{L(X, \theta) + \|\lambda \odot \theta\|_1\}}_{\text{Optimum of LASSO}} + \log \int e^{-\min_{\theta} \{L(X, \theta) + \|\lambda \odot \theta\|_1\}} dX$$

Optimum of LASSO

Complexity of LASSO

1. Optimize λ :

$$\bar{\lambda} = \underset{\lambda \geq 0}{\operatorname{argmin}} \text{RSC}(X; \lambda)$$

2. Solve LASSO with $\bar{\lambda}$:

$$\bar{\theta} = \underset{\theta \in \mathbb{R}^d}{\operatorname{argmin}} \left\{ L(X, \theta) + \|\bar{\lambda} \odot \theta\|_1 \right\}$$

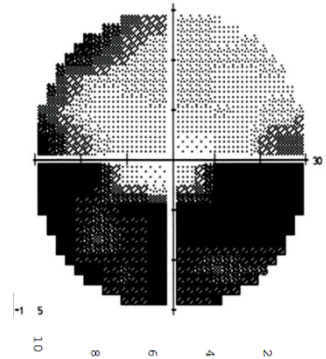
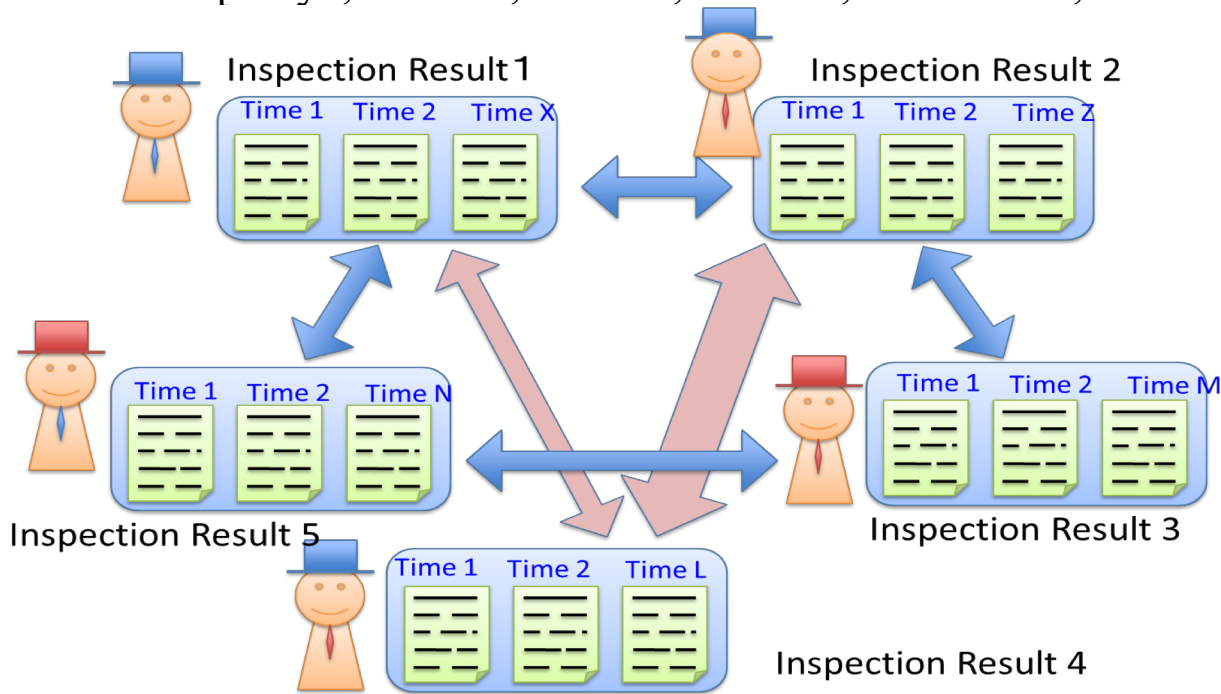
緑内障進行予測: 時空間パターン転移学習

病状の時空間パタンの似ている患者データを用いて予測精度増強

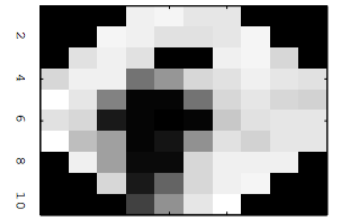
[Liang, Tomioka, Murata, Asaoka, Yamanishi ICDM013]

[Maya, Morino, Yamanishi BigData 2014]

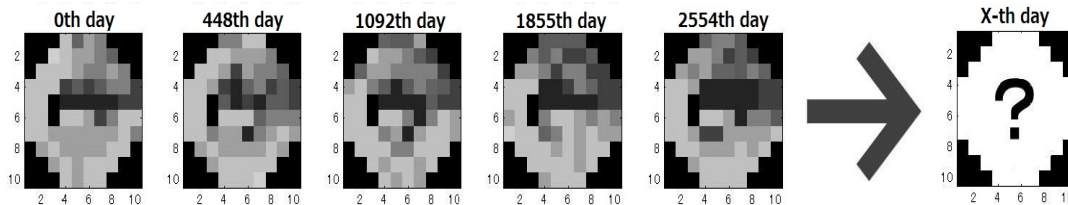
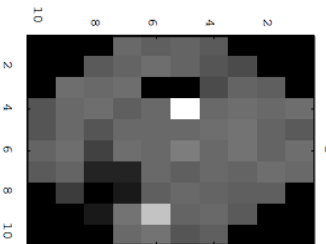
[Maya, Morino, Murata, Asaoka, Yamanishi, KDD2015]



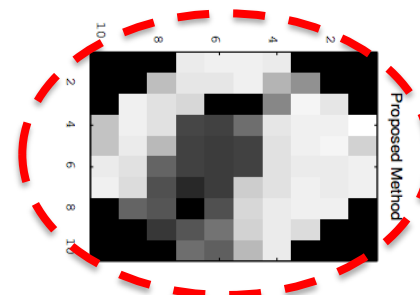
Correct



Prediction without using relational data



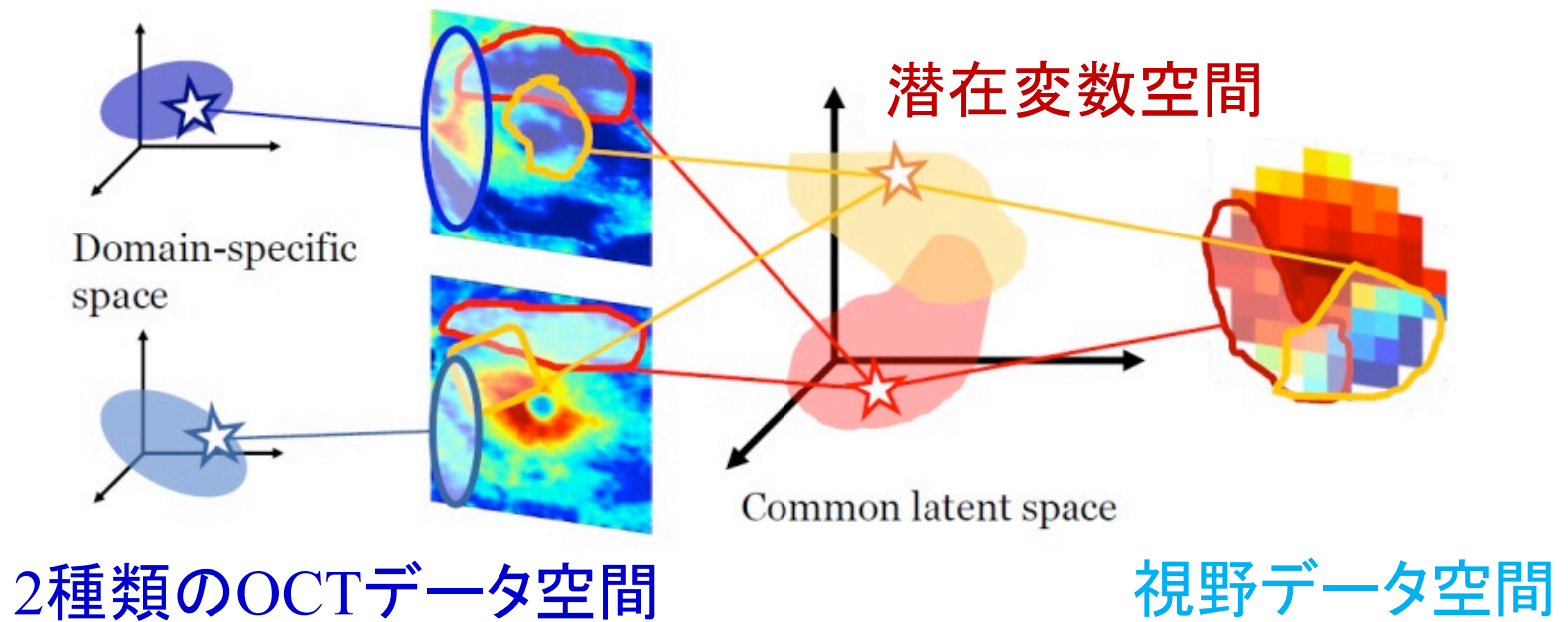
Prediction aggregating relational data



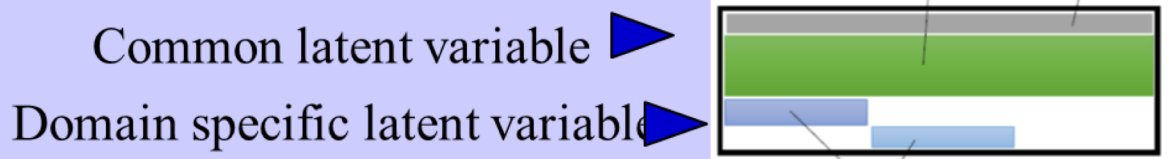
緑内障進行予測: ASNMF

測定容易なOCT(網膜厚)データを視野データに変換する技術を確立
⇒緑内障進行予測をより低コストで実現

[Uesaka, Morino, Sugiura, Kiwaki, Murata, Asaoka, Yamanishi KDD2017]



Affine Structured Nonnegative Matrix Factorization (ASNMF)



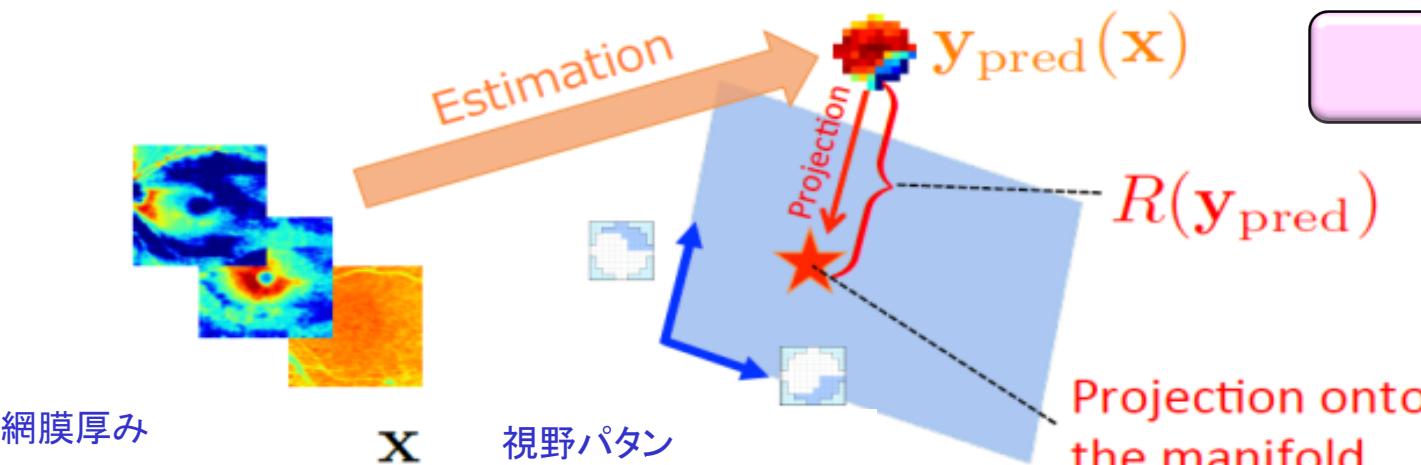
$$\min_{W, Z} \|(F - WZ) \odot \bar{M}\|_2^2$$

s. t. $W \geq 0$ and $Z \geq 0$.

緑内障進行予測：パターン正則化

パターン正則化により網膜厚みから視野感度の推定精度を大幅増強

[Sugiura, Kiwaki, Siamak, Murata, Asaoka, Yamanishi KDD2018]



パターン正則化

教師なしで学習した視野パタンからずれないように正則化

平均二乗誤差

	Median	Mean	Worst
CNN [30]	6.68	6.76	21.7
CNN + DA	6.70	6.52	21.7
CNN + DA + PBR1(NMF)	6.41	6.30	23.5
CNN + DA + PBR1(PCA)	6.31	6.26	24.1
CNN + DA + PBR1(AE)	6.39	6.26	24.2
CNN + DA + PBR1(PCA) + PBR2	6.00	6.16	21.4
CNN + DA + PBR1(AE) + PBR2	6.02	6.16	21.5

提案手法の最新結果

$$\frac{1}{N} \sum_{n=1}^N \|\mathbf{y}_{\text{pred}}(\mathbf{x}_n) - \mathbf{y}_n\|^2 + \frac{\beta}{2} R(\mathbf{y}_{\text{pred}}(\mathbf{x}_n)),$$

$$\mathbf{y}_{\text{proj}} \stackrel{\text{def}}{=} \Theta^T (\Theta \Theta^T)^{-1} \Theta (\mathbf{y}_{\text{pred}} - \mathbf{b}) + \mathbf{b}.$$

眼科学データマイニング

現世代: 視野データを用いた緑内障進行予測

クラスタリングに基づく方法

[Liang et al. ICDM2013]

変分ベイズに基づく方法

[Murata et al. IVP2014]

マルチタスク学習に基づく方法

[Maya et al. BigData2014]

緑内障因子解析

MDL原理に基づく
眼圧区分線形回帰

[Tomoda et al. Healthinf2016]

世界的に
率先する
分野

緑内障パターン発見

MDL原理に基づく
階層的クラスタリング

[Maya et al. KDD2015]

特許出願

次世代: OCTデータを用いた緑内障進行予測

特許出願

マルチビュー学習に基づく
OCTデータからの視野推定

[Uesaka et al. KDD2017]

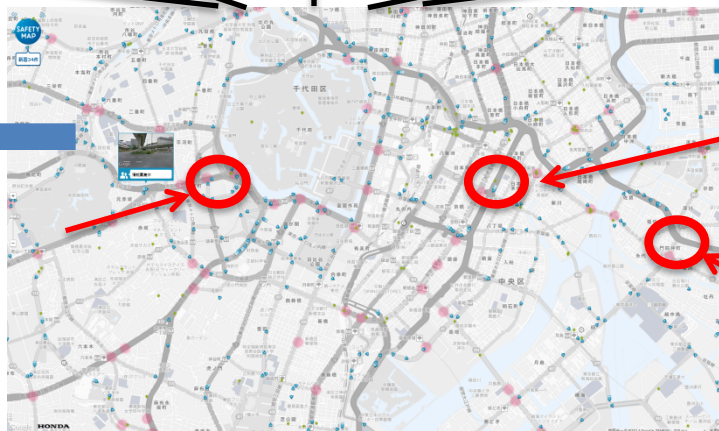
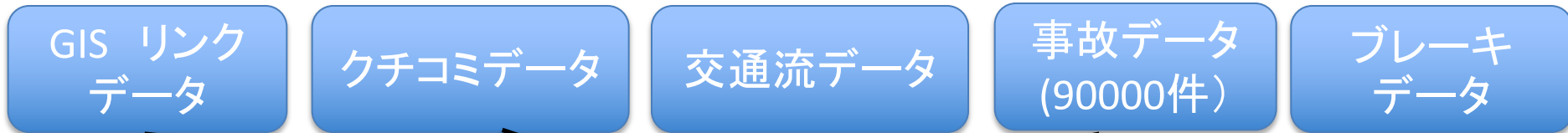
パターン正則化に基づく
OCTデータからの視野推定

[Sugiura et al. KDD2018]

交通リスクマイニング：教師なし学習

ヘテロ情報から危険道路のパターンを抽出

[Moriya, Matsushima, Yamanishi DSAA2015], [Lee, Matsushima, Yamanishi DSAA2016]



クラスター1
クラスター2
クラスター3

クラスター1
クラスター2
クラスター3

クラスター1
クラスター2
クラスター3

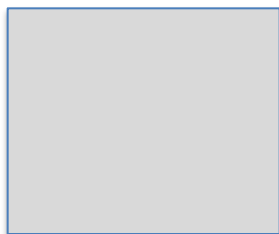
特徴ベースNMF
部分順序付きNMF

ブレ 事故 交通
ーキ 頻度 流

クラス クラス クラス
ター1 ター2 ター3

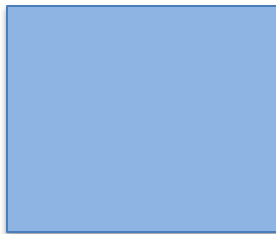
ブレ 事故 交通
ーキ 頻度 流

地点A
地点B
地点C
.....

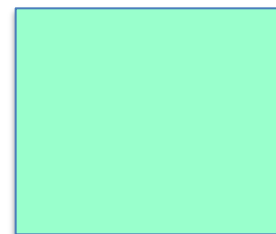


=

地点A
地点B
地点C
.....



×



クラスター1
クラスター2
クラスター3

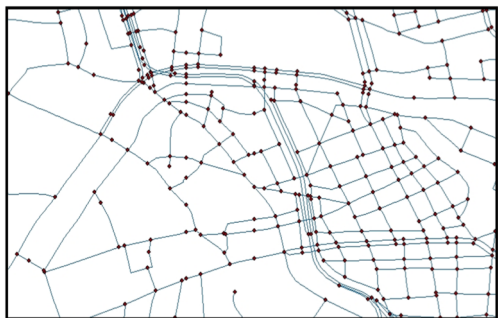
+ a + b + c

交通リスクマイニング：教師あり学習

地形データをネットワーク中心性で数値化、危険パターンを特徴づけ

[Kobayashi, Matsushima, Yamanishi BigData 2017]

地形
データ



高速パターンマイニングに基づく分類

$$\min_w C \sum_{i=1}^m L \left(\sum_{\phi \in \Phi(k)} w_{\phi} \phi(\mathbf{x}_i), y_i \right) + \|w\|_1$$

ただし、 $L(f(\mathbf{x}), y) = \log(1 + \exp(-yf(\mathbf{x})))$



Network中心性
で数値表現

(次数、近傍ノード数、
ページランク、近傍中心
性、平均リンク数 etc)



低コストデータ



- ・危険パターンの抽出
- ・潜在的危険個所の特定

SNSマイニング: 新規話題検知

リンク情報に基づくTwitterからの話題出現検知

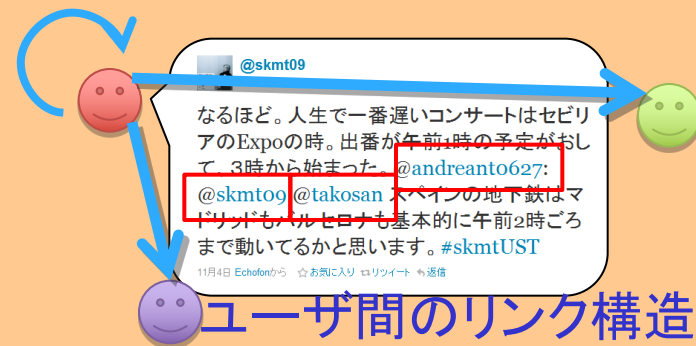
[Takahashi, Tomioka, Yamanishi ICDM2011, TKDE2014]

従来の話題検出法

GW 原発
母の日
単語に注目

問題点: 言い換えに弱い

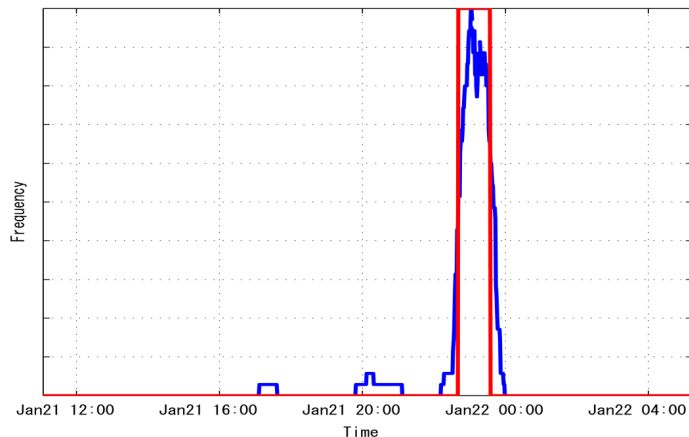
提案する話題検出法



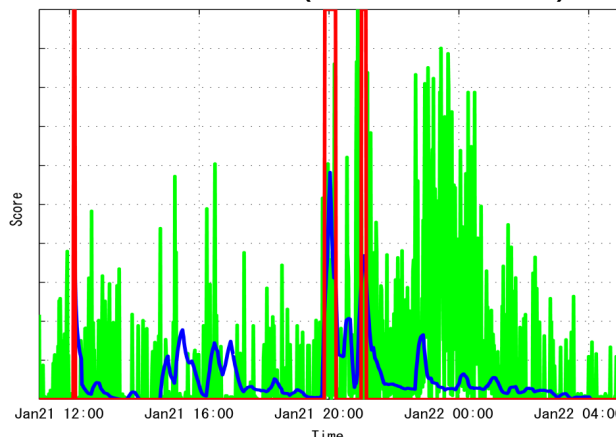
ユーザ間のリンク構造の
変化検知により話題検知

従来手法(キーワード頻度)より速い話題検出が可能

比較手法(キーワード頻度)



提案手法(変化点検出)



SNSマイニング: 持続性話題検知

リンクネットワーク構造の異常検知により長期間流行話題を検知

[Saito, Tomioka, Yamanishi ASONAM2014,]

長期間流行するTweet

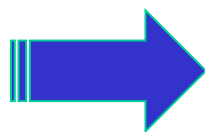
 @chibicode
上杉周作

アップルで働くまで、イノベーションというのは「今にない、新しいものを作ること」だと思っていた。でもそれは違って、イノベーションというのは「未来にある普通のものを作ること」なのです。この違いを理解できるまでかなり時間がかかった。

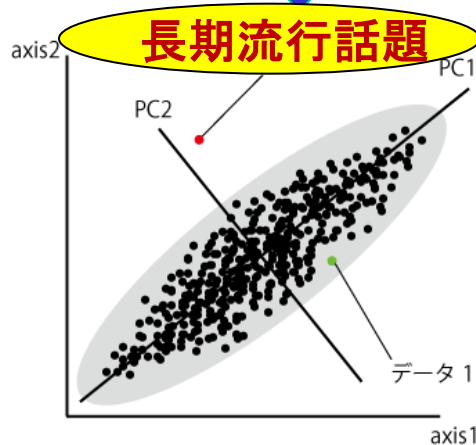
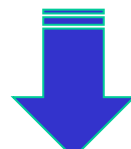
2月5日 Echofonから

☆ お気に入り登録 リツイート 返信

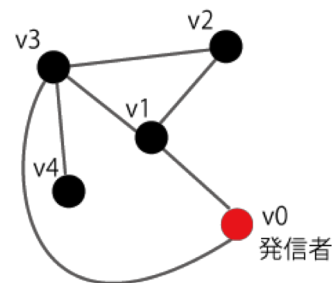
shimo3と他100+人がリツイート



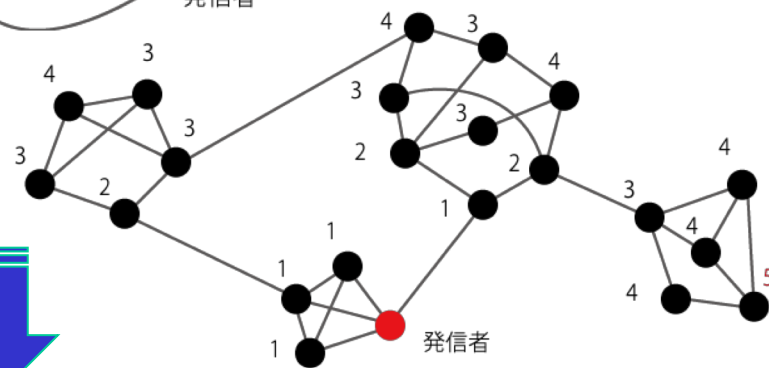
グラフ的な性質
で特徴づけ



共有者数



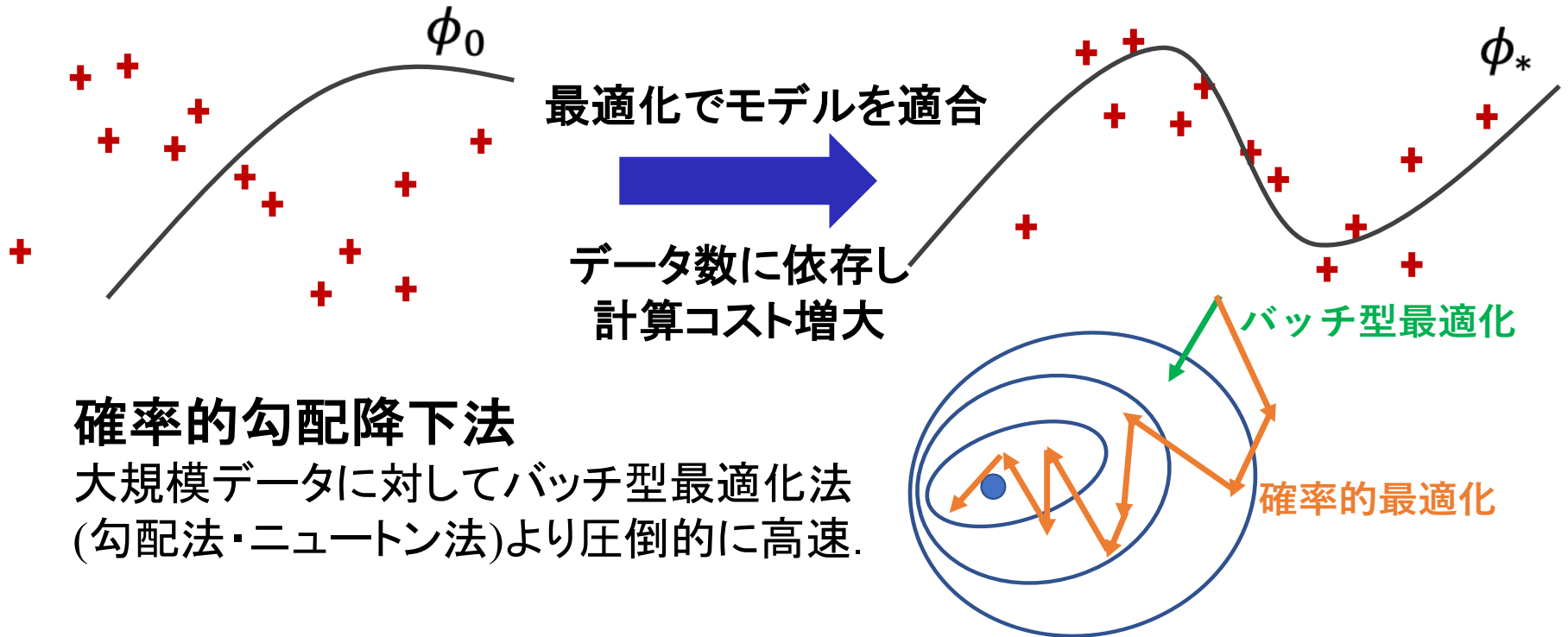
最大距離



異常検知

確率的最適化

大規模機械学習用の高速最適化手法



問題固有の構造を利用した更なる高速化の研究

- 1: **加速法＋分散縮小法** [Nitanda NIPS2014, AISTATS2016]
最適反復計算量を最小ミニバッチサイズで達成した世界初の手法.
- 2: **確率的DC最適化法** [Nitanda, Suzuki AISTATS2017]
DC構造を利用しボルツマンマシンのEMアルゴリズムを改良・高速化.

関数勾配法による深層ニューラルネット学習

関数勾配を用いたResNetの理論解析と手法開発

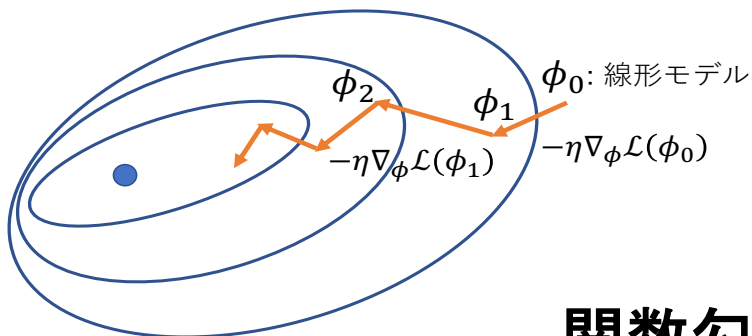
ResNet: 残差ブロックを待つ深層ニューラルネット

関数勾配法的解釈

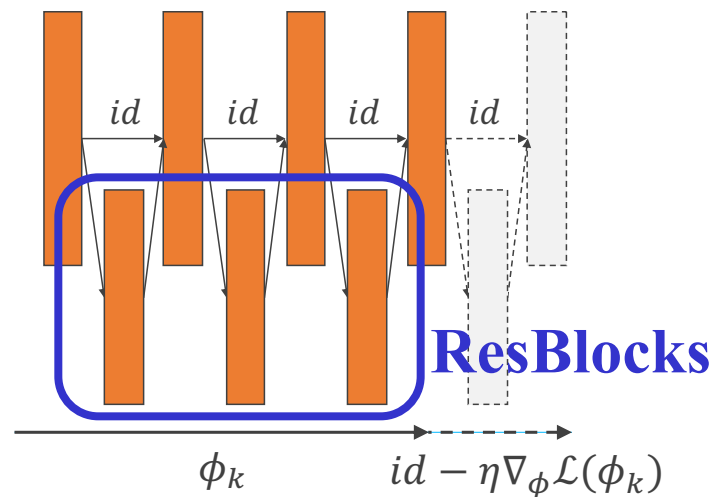
[Nitanda, Suzuki AISTATS2018, ICML2018]

$$\phi_{s+1} = (id + F_s) \circ \dots \circ (id + F_0) \circ id$$

$$-\eta \nabla_{\phi} \mathcal{L}(\phi_t) \sim F_t \circ \phi_t = \phi_{t+1} - \phi_t$$



層が積み重なる過程を
関数最適化法と見なす。



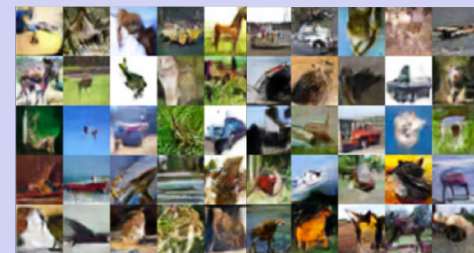
関数勾配法の応用例

1: **生成モデル学習** [Nitanda, Suzuki AISTATS2018]

輸送理論を用いた収束解析。

2: **識別学習** [Nitanda, Suzuki ICML2018]

ブースティング理論を用いた汎化解析。



2015－2016の修論テーマ

- ・ネットワーク中心性を用いたテンポラルネットワークの変化検知
DSAA16採択
- ・ヘテロデータに基づく緑内障進行予測のための個々の患者に特化した病状のモデル化
Healthinf16採択 2015研究科長賞(創造)
- ・記述長最小化原理に基づく関係データ学習の研究
SDM16採択
- ・Support Vector Machine 及びトピックモデルを用いた位置情報による個人識別の研究
- ・完全変数化NML 符号長に基づく正準相関分析のランク推定の研究
DSAA17採択
- ・Study on Learning from Nonstationary Time Series
DSAA15採択
- ・サポートベクトルマシンの最適化アルゴリズムの研究
- ・野球データ解析に対する機械学習アプローチ
DSAA16採択
- ・交通データからのリスク因子発見に関する研究
2016研究科長賞(創造)
- ・緑内障眼における網膜層厚視野感度間マルチビュー学習
KDD17採択
- ・A Study on Model Selection for Hierarchical Mixture Models
2016研究科長賞(数理)
- ・畳み込み行列分解のパラメータ推定と構造選択
KDD17採択
ICDM16採択

2017の修論テーマ

- Model Selection for Non-negative Tensor Factorization with Minimum Description Length
- パターン正則化学習を用いた網膜層厚からの視野感度推定と知識発見

KDD18採択

2016-2017の卒論テーマ

- 可変窓上の記述長最小化に基づく時系列データの変化検知
- ネットワーク情報を用いた道路の潜在的危険予測
- オンライン非定常ロバスト学習とその異常検知への応用
- ネットワーク情報を用いた道路の潜在的危険予測

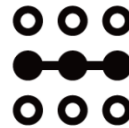
BigData17採択

BigData17採択

2017工学部長賞(計数)

東京大学大学院情報理工学系研究科 領域知識創成教育研究プログラム (東京大学データサイエンティスト養成講座) の主担当協力研究室

最先端の情報理工学で
データサイエンスの未来を
切り開く。



DSS

Data Science School

情報理工学の基礎を学び、ビジネスデータ解析で応用力を鍛え、最先端科学へ挑戦する。
東京大学データサイエンティスト養成講座はじまります。

受講生募集

受講の応募ごとに東京大学の電子証明書を発行いたします。

<http://dss.i.u-tokyo.ac.jp>



数理6研はDSSをサポートしています