

# 機械学習のこれから：汎用的なデータ解析を目指して

## Machine Learning in Future: Towards Versatile Data Analysis

杉山将

Masashi Sugiyama

東京大学 複雑理工学専攻

Department of Complexity Science and Engineering, The University of Tokyo

sugi@k.u-tokyo.ac.jp <http://www.ms.k.u-tokyo.ac.jp>

産業界や基礎科学の様々な分野において、大量のデータの山から新たな価値を創造する機械学習技術の重要性が増している。しかし、解析すべきデータの量・次元・複雑さが爆発的に増加しているため、データ解析手法の研究・開発が社会的なニーズに追いつかなくなりつつある。また、最先端のデータ解析アルゴリズムは極めて高度な確率論・統計学・最適化理論等を駆使して設計されているため、技術修得が著しく困難であるという問題もある。

そこで我々は、データ解析に「データ解析コア技術」という独自の抽象的な階層を導入することを提案している。これは、分類、回帰、特徴選択、異常検出などの主要なデータ解析タスクからなる部分集合を考えるものであり、これらのタスク群に共通して適用できるデータ解析基盤技術を開発してきた。本講演では、確率分布間の距離の推定や情報量などを用いた汎用的な機械学習技術を紹介するとともに、それらの応用例や最新の研究成果についても述べる。

### 参考文献

- [1] 杉山 将. 密度比推定によるビッグデータ解析. 電子情報通信学会誌, vol.97, no.5, pp.353-358, 2014. <http://www.ms.k.u-tokyo.ac.jp/2014/IEICE-DensityRatioReview-jp.pdf>
- [2] 杉山 将. 確率分布間の距離推定：機械学習分野における最新動向. 日本応用数理学会論文誌, vol.23, no.3, pp.439-452, 2013. <http://www.ms.k.u-tokyo.ac.jp/2013/DivergenceReview-jp.pdf>
- [3] Sugiyama, M., Suzuki, T., & Kanamori, T. Density Ratio Estimation in Machine Learning, Cambridge University Press, Cambridge, UK, 2012.

## 機械学習

1

- **機械学習**: データの背後に潜む知識を学習する
- **様々な応用例**:
  - 音声・画像・動画の認識
  - ウェブやSNSからの情報抽出
  - 商品やサービスの推薦
  - 工業製品の品質管理
  - ロボットシステムの制御
- **ビッグデータ**時代の到来に伴い、機械学習技術の重要性は益々高まりつつある

## 機械学習のタスク

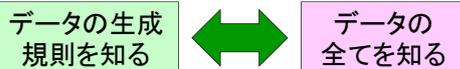
2

- **機械学習**には様々なタスクがある:
  - 非定常環境下での適応学習, ドメイン適応, マルチタスク学習
  - 二標本検定, 異常値検出, 変化点検知, クラスバランス推定
  - 相互情報量推定, 独立性検定, 特徴選択, 十分次元削減, 独立成分分析, 因果推論, クラスタリング, オブジェクト適合
  - 条件付き確率推定, 確率的パターン認識

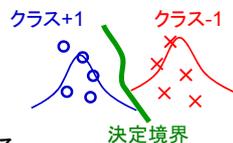
## 最も汎用的なアプローチ

3

- **データを生成する規則(確率分布)**を推定すれば、あらゆる機械学習タスクが解決できる!



- 例: 各クラスのデータの生成分布がわかれば、パターン認識ができる



- **生成的アプローチ**とよばれる

## 各タスクに特化したアプローチ

4

- しかし、**確率分布の推定は困難**であるため、生成モデル推定に基づくアプローチによって、必ずしも高い学習精度が得られるとは限らない

- 確率分布の推定を行わず、各タスクを直接解く

- 例: サポートベクトルマシンでは、各クラスのデータ生成分布を推定せず、パターン認識に必要な決定境界のみを学習

- パターン認識に対しては、**識別的アプローチ**とよばれる



## 各タスクに特化したアプローチ

5

- 各タスクに特化したアルゴリズムを開発した方が原理的には生成的アプローチよりも性能が良い

- しかし、様々なタスクに対して**個別に**研究開発を行うのは大変:

- アルゴリズム考案
- 理論的性能評価
- 高速かつメモリ効率の良い実装
- エンジニアの技術習得

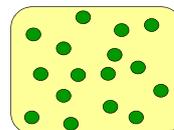
## 本日紹介するアプローチ

6

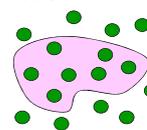
- **中間的なアプローチ**: あるクラスのタスク群に対して、研究開発を行う

- 確率密度比, 確率密度差, 距離, 情報量, 確率密度微分などの抽象的な量の推定を通して、データ解析を行う

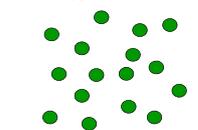
生成的アプローチ



中間アプローチ



タスク特化アプローチ



## 確率密度比に基づく機械学習 7

- 前述の機械学習タスク群は**複数の確率分布**を含む

$$p(\mathbf{x}), q(\mathbf{x})$$

- しかし、これらのタスクを解くのに、それぞれの確率分布そのものは**必要ない**
- 確率密度関数の**比**が分かれば十分である

$$r(\mathbf{x}) = \frac{p(\mathbf{x})}{q(\mathbf{x})}$$

- 各確率分布は推定せず、**密度比を直接推定**することにする

## 直感的な正当化 8

### パニックの原理 Vapnik (1998)

ある問題を解くとき、それより一般的な問題を途中段階で解くべきでない

$p(\mathbf{x}), q(\mathbf{x})$   
が分かる



$r(\mathbf{x}) = \frac{p(\mathbf{x})}{q(\mathbf{x})}$   
が分かる

- **密度を求めるよりも、密度比を求めるほうが易しい**



Sugiyama, Suzuki & Kanamori,  
Density Ratio Estimation  
in Machine Learning,  
Cambridge University Press, 2012



## 発表の流れ 9

1. 密度比推定に基づく機械学習の枠組み
2. **密度比推定法**
3. 密度比推定の応用事例
4. 発展的な話題

## 最小二乗密度比適合 10

Kanamori, Hido & Sugiyama (JMLR2009)

- データ:  $\{\mathbf{x}_i^p\}_{i=1}^{n_p} \overset{i.i.d.}{\sim} p(\mathbf{x}), \{\mathbf{x}_j^q\}_{j=1}^{n_q} \overset{i.i.d.}{\sim} q(\mathbf{x})$
- 真の密度比  $r(\mathbf{x})$  との **二乗誤差** を最小にするように密度比モデル  $r_\alpha(\mathbf{x})$  を学習:

$$\begin{aligned} J(\alpha) &= \frac{1}{2} \int (r_\alpha(\mathbf{x}) - r(\mathbf{x}))^2 q(\mathbf{x}) d\mathbf{x} & r(\mathbf{x}) &= \frac{p(\mathbf{x})}{q(\mathbf{x})} \\ &= \frac{1}{2} \int r_\alpha(\mathbf{x})^2 q(\mathbf{x}) d\mathbf{x} - \int r_\alpha(\mathbf{x}) p(\mathbf{x}) d\mathbf{x} + C \\ &\approx \frac{1}{2n_q} \sum_{j=1}^{n_q} r_\alpha(\mathbf{x}_j^q)^2 - \frac{1}{n_p} \sum_{i=1}^{n_p} r_\alpha(\mathbf{x}_i^p) + C \end{aligned}$$

## アルゴリズム 11

- 密度比モデル:  $r_\alpha(\mathbf{x}) = \sum_{\ell=1}^{n_p} \alpha_\ell \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}_\ell^p\|^2}{2\sigma^2}\right)$

- 最適化規準:  $\min_{\alpha} \left[ \frac{1}{2} \alpha^\top \hat{G} \alpha - \hat{\mathbf{h}}^\top \alpha + \frac{\lambda}{2} \alpha^\top \alpha \right]$

$$\hat{G}_{\ell, \ell'} = \frac{1}{n_q} \sum_{j=1}^{n_q} \exp\left(-\frac{\|\mathbf{x}_j^q - \mathbf{x}_\ell^p\|^2}{2\sigma^2}\right) \exp\left(-\frac{\|\mathbf{x}_j^q - \mathbf{x}_{\ell'}^p\|^2}{2\sigma^2}\right)$$

$$\hat{\mathbf{h}}_\ell = \frac{1}{n_p} \sum_{i=1}^{n_p} \exp\left(-\frac{\|\mathbf{x}_i^p - \mathbf{x}_\ell^p\|^2}{2\sigma^2}\right)$$

- 大域的最適解が解析的に計算可能:

$$\hat{\alpha} = (\hat{G} + \lambda I)^{-1} \hat{\mathbf{h}}$$



## 最小二乗密度比適合の 12

### MATLABによる実装

$$\hat{\alpha} = (\hat{G} + \lambda I)^{-1} \hat{\mathbf{h}} \quad \hat{G}_{\ell, \ell'} = \frac{1}{n_q} \sum_{j=1}^{n_q} \exp\left(-\frac{\|\mathbf{x}_j^q - \mathbf{x}_\ell^p\|^2}{2\sigma^2}\right) \exp\left(-\frac{\|\mathbf{x}_j^q - \mathbf{x}_{\ell'}^p\|^2}{2\sigma^2}\right)$$

$$\hat{\mathbf{h}}_\ell = \frac{1}{n_p} \sum_{i=1}^{n_p} \exp\left(-\frac{\|\mathbf{x}_i^p - \mathbf{x}_\ell^p\|^2}{2\sigma^2}\right)$$

#### %人工データの生成

```
n=300; x=randn(n,1); y=randn(n,1)+0.5;
```

#### %密度比の推定

```
x2=x.^2; xx= repmat(x2,1,n)+ repmat(x2',n,1)-2*x*x';
y2=y.^2; yx= repmat(y2,1,n)+ repmat(x2',n,1)-2*y*x';
r=exp(-yx); s=r*((r+eye(n))\mean(exp(-xx),2)); plot(y,s,'rx');
```

## 理論解析

13

■ **パラメトリックモデルの場合:**  $r_{\alpha}(x) = \sum_{\ell=1}^b \alpha_{\ell} \phi_{\ell}(x)$

- 学習したパラメータは  $n^{-\frac{1}{2}}$  の速さで最適値に収束
- 最適な収束率を達成している  $n = \min(n_p, n_q)$

Kanamori, Hido & Sugiyama (JMLR2009)

■ **ノンパラメトリックモデルの場合:**  $r_{\alpha}(x) = \sum_{\ell=1}^{n_p} \alpha_{\ell} \exp\left(-\frac{\|x - x_{\ell}^p\|^2}{2\sigma^2}\right)$

- 学習した関数は  $n^{-\frac{1}{2+\gamma}}$  の速さで真の関数に収束 (関数空間のブラケットエントロピーに依存)
- 最適な収束率を達成している  $0 < \gamma < 2$

Kanamori, Suzuki & Sugiyama (ML2012)



## 発表の流れ

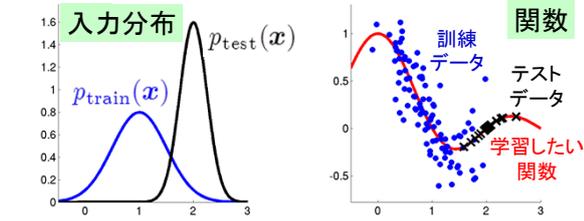
14

1. 密度比推定に基づく機械学習の枠組み
2. 密度比推定法
3. 密度比推定の応用事例
4. 発展的な話題

## 共変量シフト適応

15

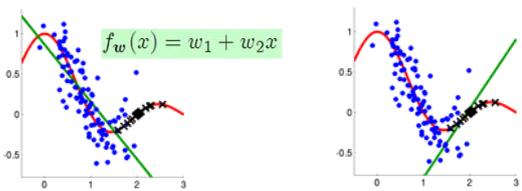
- 共変量とは入力変数の別名
- **共変量シフト:** 訓練時とテスト時で入力分布が変化するが、入出力関数は変わらない
- **外挿問題**が典型的な例



## 重要度重み付き最小二乗学習

16

$$\min_w \sum_{i=1}^n (f_w(x_i) - y_i)^2 \quad \min_w \sum_{i=1}^n \frac{p_{\text{test}}(x_i)}{p_{\text{train}}(x_i)} (f_w(x_i) - y_i)^2$$



- 共変量シフト下では、通常の最小二乗学習は一般性を欠く ( $n \rightarrow \infty$  でも最適解に収束しない)
- 共変量シフト下でも一般性を持つ様々な学習法に適用可能:
  - サポートベクトルマシン, ロジスティック回帰, 条件付き確率場など

## 実世界応用例

17

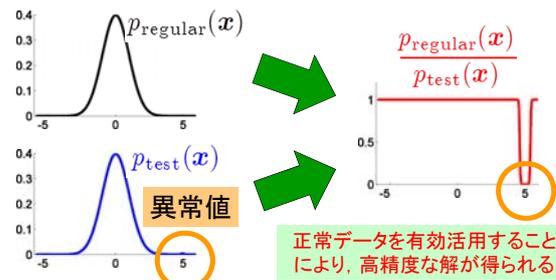
- **顔画像からの年齢予測:**
  - 照明環境の変化 (Ueki, Sugiyama & Ihara (IEICE-ED2011))
- **話者認識:**
  - 声質の変化 (Yamada, Sugiyama & Matsui (SigPro2010))
- **テキスト分割:**
  - ドメイン適応 (Tsuboi, Kashima, Hido, Bickel & Sugiyama (JIP2008))
- **ブレイン・コンピュータインターフェース:**
  - 心理状態の変化 (Sugiyama, Krauledat & Müller (JMLR2007), Li, Kambara, Koike & Sugiyama (IEEE-TBE2010))

## 正常値に基づく異常値検出

18

Hido, Tsuboi, Kashima, Sugiyama & Kanamori (KAIS2011)

- 正常データと傾向が異なるテストデータを異常値とみなす。



## 実世界応用例

19

### ■ 製鉄プロセスの異常診断

Hirata, Kawahara & Sugiyama (Patent2010)

### ■ 光学部品の品質検査

Takimoto, Matsugu & Sugiyama (DMSS2009)

### ■ ローン顧客の審査

Hido, Tsuboi, Kashima, Sugiyama & Kanamori (KAIS2011)

## 二標本検定

20

Sugiyama, Suzuki, Ito, Kanamori & Kimura (NN2011)

- **目的:** 二つのデータセットの背後の確率分布が同じかどうかを検定する

$$\{x_i^p\}_{i=1}^{n_p} \overset{i.i.d.}{\sim} p(x)$$

$$\{x_j^q\}_{j=1}^{n_q} \overset{i.i.d.}{\sim} q(x)$$

- **アプローチ:** 密度比を用いて分布間の距離を推定する

- カルバック・ライブラー距離:  $\int p(x) \log \frac{p(x)}{q(x)} dx$

- ピアソン距離:  $\int q(x) \left( \frac{p(x)}{q(x)} - 1 \right)^2 dx$

## 実世界応用例

21

### ■ 画像中の注目領域抽出

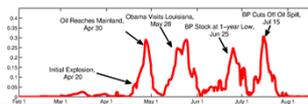
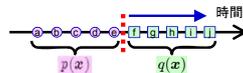
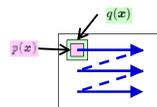
Yamanaka, Matsugu & Sugiyama (IPSJ-TOM2013)

### ■ 動画からのイベント検出

Yamanaka, Matsugu & Sugiyama (IPSJ-TOM2013)

### ■ ツイッターデータ解析

Liu, Yamada & Sugiyama (NN2013)



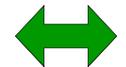
## 相互情報量推定

22

Suzuki, Sugiyama, Sese & Kanamori (FSDM2008), Sugiyama (Entropy2013)

- **相互情報量:**  $MI = \int p(x, y) \log \frac{p(x, y)}{p(x)p(y)} dx dy$

MI = 0



$x$  と  $y$  は  
統計的に独立

- 相互情報量は**密度比**を用いて計算できる
- 最小二乗密度比推定には、  
二乗損失相互情報量が自然:

$$SMI = \int p(x)p(y) \left( \frac{p(x, y)}{p(x)p(y)} - 1 \right)^2 dx dy$$

## 相互情報量に基づく機械学習

23

### ■ 入出力間の独立性判定:

- 特徴選択
- クラスタリング

Suzuki, Sugiyama, Sese & Kanamori (BMC-Bioinfo2009)

Suzuki & Sugiyama (NeCo2012)

Sugiyama, Niu, Yamada, Kimura & Hachiya (NeCo2013)

### ■ 実世界応用例:

- 遺伝子解析
- 画像認識
- 音響認識



## 相互情報量に基づく機械学習

24

### ■ 入力間の独立性判定:

- 独立成分分析
- オブジェクト適合

Suzuki & Sugiyama (NeCo2011)

Karasuyama & Sugiyama (NN2012)

Yamada & Sugiyama (AISTATS2011)

### ■ 実世界応用例:

- モーションキャプチャデータの解析
- 医療画像の位置合わせ
- 写真の自動レイアウト



## 条件付き確率密度の推定 25

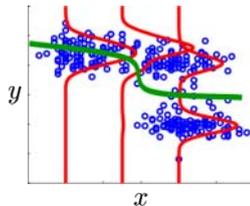
$$p(y|x) = \frac{p(x, y)}{p(x)}$$

Sugiyama, Takeuchi, Suzuki, Kanamori, Hachiya & Okanohara (IEICE-ED2010)

- 回帰分析: 条件付き期待値の推定
- 非対称なノイズや多峰性を持つようなデータに対しては、回帰分析では不十分
- 実世界応用例:

- ヒューマノイドロボット制御

Sugimoto, Tangkaratt, Wensveen, Zhao, Sugiyama & Morimoto (HUMANOIDS2014)

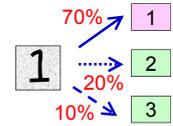


## 確率的パターン認識 26

$$p(y|x) = \frac{p(x, y)}{p(x)}$$

Sugiyama (IEICE-ED2010)

- 出力  $y$  がカテゴリのとき、条件付き確率の推定は確率的なパターン認識に対応



- 実世界応用例:

- 顔画像からの年齢推定  
Ueki, Sugiyama, Ihara & Fujita (ACPR2011)

- 加速度データからの行動認識  
Hachiya, Sugiyama & Ueda (Neurocomputing2012)

## 発表の流れ 27



1. 密度比推定に基づく機械学習の枠組み
2. 密度比推定法
3. 密度比推定の応用事例
4. 発展的な話題

## 発展的な話題 28

- ブレグマン距離を用いた密度比推定の統一理論

Sugiyama, Suzuki & Kanamori (AISM2012)

- 次元削減付き密度比推定

Sugiyama, Kawanabe & Chui (NN2010)  
Sugiyama, Yamada, von Büna, Suzuki, Kanamori & Kawanabe (NN2011)

- 相対密度比推定

Yamada, Suzuki, Kanamori, Hachiya & Sugiyama (NIPS2011, NeCo2013)

$$\frac{p(x)}{\beta p(x) + (1 - \beta)q(x)} < \frac{1}{\beta} \quad 0 < \beta < 1$$

- 密度差推定

Sugiyama, Suzuki, Kanamori, du Plessis, Liu & Takeuchi (NIPS2012, NeCo2013)

$$p(x) - q(x)$$

## 密度比の世界 29

### 実問題応用例:

ブレイン・コンピュータインターフェース、ロボット制御、音声認識、画像認識、自然言語処理、バイオインフォマティクス、データマイニング

### 機械学習アルゴリズム:

重点サンプリング(共変量ソフト適応, ドメイン適応, 多タスク学習), 二標本問題(二標本検定, 外れ値検出, 変化点検知), 相互情報量推定(独立性検定, 変数選択, 独立成分分析, 次元削減, 因果推定, クラスタリング, オブジェクト適合), 条件付き確率推定(可視化, 状態遷移推定, 確率的パターン認識),

### 密度比推定法:

基本アルゴリズム(LR, KMM, KLIEP, LSIF), 大規模対応, 高次元対応, 安定化, ロバスト化, 統一化

### 理論解析:

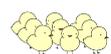
収束性解析(確率論), 情報量規準(統計学), 安定性解析(最適化)

## まとめ 30

- 密度比は、単純な最小二乗法で精度・効率良く推定できる

- 多くの学習タスクが実は最小二乗法で解ける:

- 重点サンプリング:  $\sum_{i=1}^n \frac{p_{\text{test}}(x_i)}{p_{\text{train}}(x_i)} \text{loss}(x_i)$



- ダイバージェンス推定:  $\int p(x) \log \frac{p(x)}{q(x)} dx$



- 相互情報量推定:  $\iint p(x, y) \log \frac{p(x, y)}{p(x)p(y)} dx dy$

- 条件付き確率推定:  $p(y|x) = \frac{p(x, y)}{p(x)}$