

不完全な情報からの 機械学習

杉山 将

理化学研究所／東京大学



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



東京大学
THE UNIVERSITY OF TOKYO

自己紹介

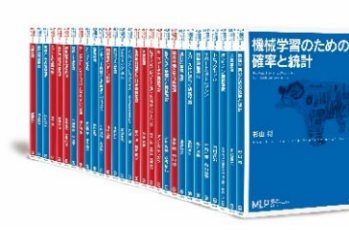
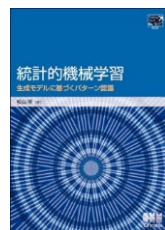
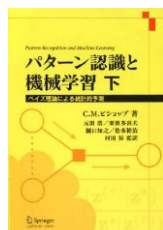
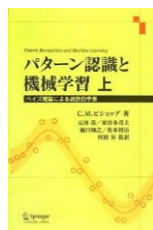
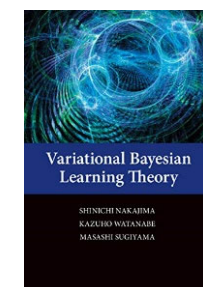
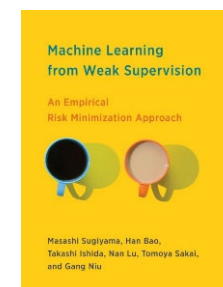
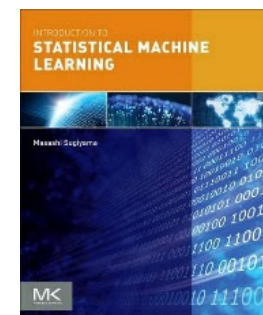
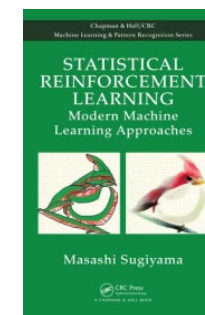
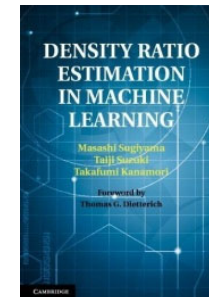
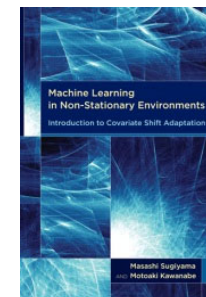
2

■ 現職:

- 理化学研究所・センター長: **研究者とともに**
- 東京大学・教授: **学生とともに**
- 企業・技術顧問: **経営者, エンジニアとともに**

■ 専門分野:

- 機械学習の数学的な基礎研究
(コンピュータ科学, 統計学など)
- 機械学習技術の実世界応用
(画像, 言語, 脳波, ロボット, 医療, 生命など)
- 執筆活動

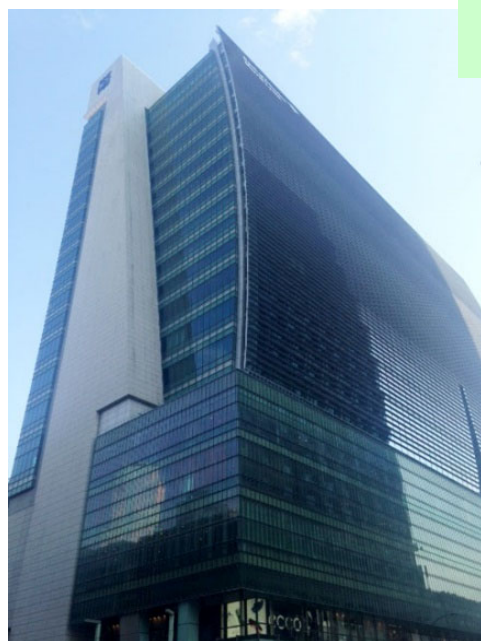


理化学研究所 革新知能統合研究(AIP)センター

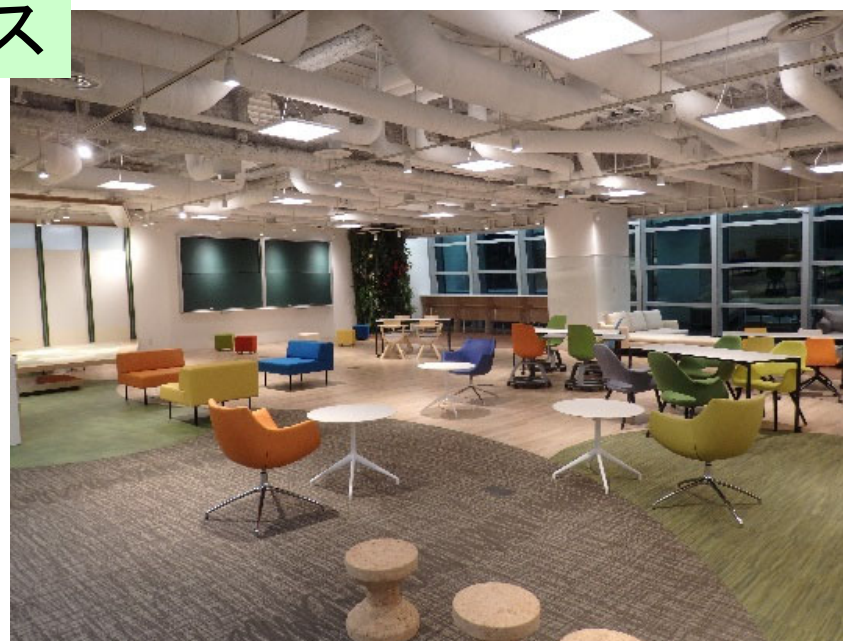
■ 文部科学省AIPプロジェクト(2016-2025年度) を推進する研究組織

- 常勤研究員137名(外国人42%, 女性23%)
- 客員研究員257名, 学生96名
- 海外インターン生延べ197名

〈2025年4月現在〉



日本橋オフィス



理研AIPの3つの研究グループ

4

汎用基盤技術 (16)

不完全情報学習チーム	構造的学習チーム
 杉山 将	 河原 吉伸
テンソル学習チーム	関数解析の学習チーム
 Qibin Zhao	 Quang Minh Ha
圧縮情報処理チーム	深層学習理論チーム
 田部井 靖生	 鈴木 大慈
計算論的学習理論チーム	因果推論チーム
 畑埜 晃平	 清水 昌平
近似ベイズ推論チーム	連続最適化チーム
 Mohammad Emtiyaz Khan	 武田 朗子
数理科学チーム	高次元因果解析チーム
 坂内 健一	 今泉 允聡
逐次的意思決定チーム	不確実性定量化チーム
 伊藤 伸志	 二見 太
情報統計力学チーム	計算的社会選択理論チーム
 坂田 綾香	 五十嵐 歩美

- 学習と推論のアルゴリズム
 - ロバスト、ベイズ、構造、...
- 統計的学習と最適化の理論
 - 収束性、表現力、非凸、...

目的指向基盤技術 (24)

がん探索医療研究チーム	iPS細胞連携医学的リスク回避チーム	分子情報科学チーム
 浜本 隆二	 上田 修功	 津田 宏治
認知行動支援技術チーム	防災科学チーム	インフラ管理ロボット技術チーム
 大武 美保子	 上田 修功	 岡谷 貴之
医用機械知能チーム	データ駆動型生物医科学チーム	計算脳ダイナミクスチーム
 原田 達也	 竹内 一郎	 山下 宙人
遺伝統計学チーム	病理情報学チーム	自然言語理解チーム
 田宮 元	 山本 陽一郎	 乾 健太郎
知識獲得チーム	空間情報学チーム	音楽情報知能チーム
 松本 裕治	 横矢 直人	 浜中 雅俊
音響情景理解チーム	計算物理機械学習チーム	統計宇宙科学チーム
 吉井 和佳	 谷口 隆晴	 吉田 直紀
化学反応情報学チーム	ロボットラーニングチーム	ロボットシステムチーム
 瀧川 一学	 長 隆之	 岡田 慧
AIコンピューティングチーム	三次元環境情報理解チーム	説明可能AIチーム
 高前田 伸也	 金崎 朝子	 谷中 瞳

- AIによる科学研究の加速
 - がん、材料、遺伝子、...
- AIによる社会課題解決への貢献
 - 自然災害、高齢者ヘルスケア、ロボット、...

社会におけるAI (6)

社会におけるAI利用と法制度チーム
 中川 裕志
科学技術と社会チーム
 佐倉 統
分散型ビッグデータチーム
 橋田 浩一
経済経営情報融合分析チーム
 星野 崇宏
人工知能セキュリティ・プライバシーチーム
 佐久間 淳
人工知能安全性・信頼性ユニット
 荒井 ひろみ

- 個人データ管理
- AIセキュリティ
- 経済と経営
- 倫理とガバナンス

<2025年5月現在>



講義の流れ

5

1. 機械学習の研究分野の紹介
2. 不完全な教師情報からの機械学習
 - A) 弱教師付き学習
 - B) ラベル雑音ロバスト学習
 - C) 転移学習
 - D) 最新の取り組み

機械学習

■ 機械学習：データの背後に潜む知識を学習する

- 人工知能の1つの研究分野

■ 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
- 医用画像処理

■ 近年，機械学習技術の重要性は益々高まっている

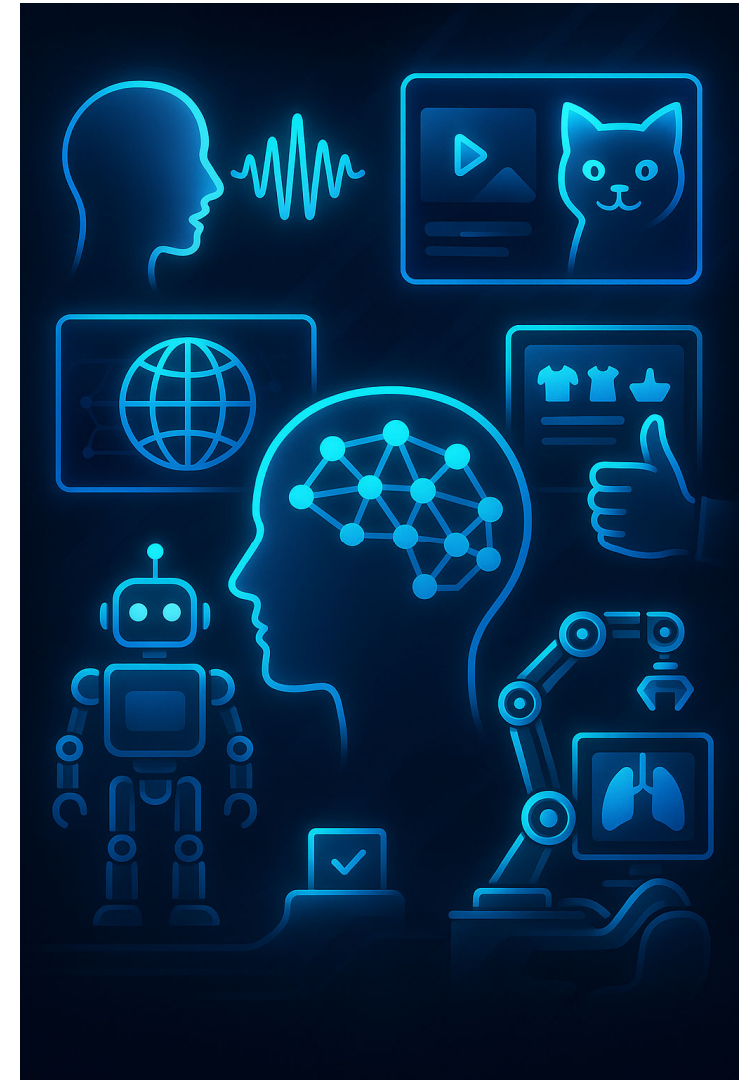


Image generated by ChatGPT 4o

人工知能・機械学習分野の歴史

7

■ 論理的知能:

- 1960年代:
記号処理, 論理推論
- 1980年代:
エキスパートシステム

■ 脳型情報処理:

- 1960年代:
パーセプトロン(1層)
- 1980年代:
誤差逆伝播法(多層)

■ 統計的機械学習:

- 2000年代: 統計・凸最適化,
カーネル法, ベイズ推論

■ 生成AI:

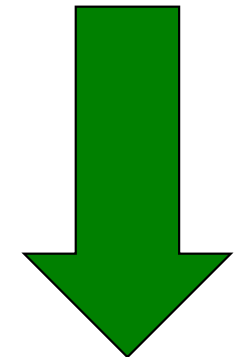
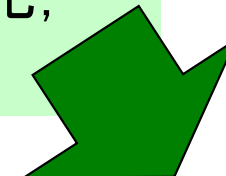
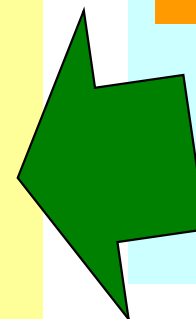
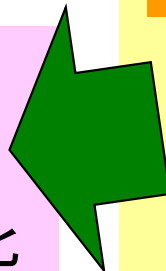
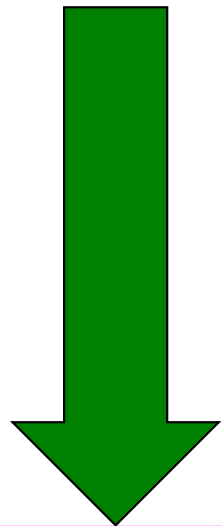
- 2020年代:
画像生成,
対話

■ 深層学習:

- 2010年代:
確率的降下法,
巨大深層モデル

■ 次世代知能:

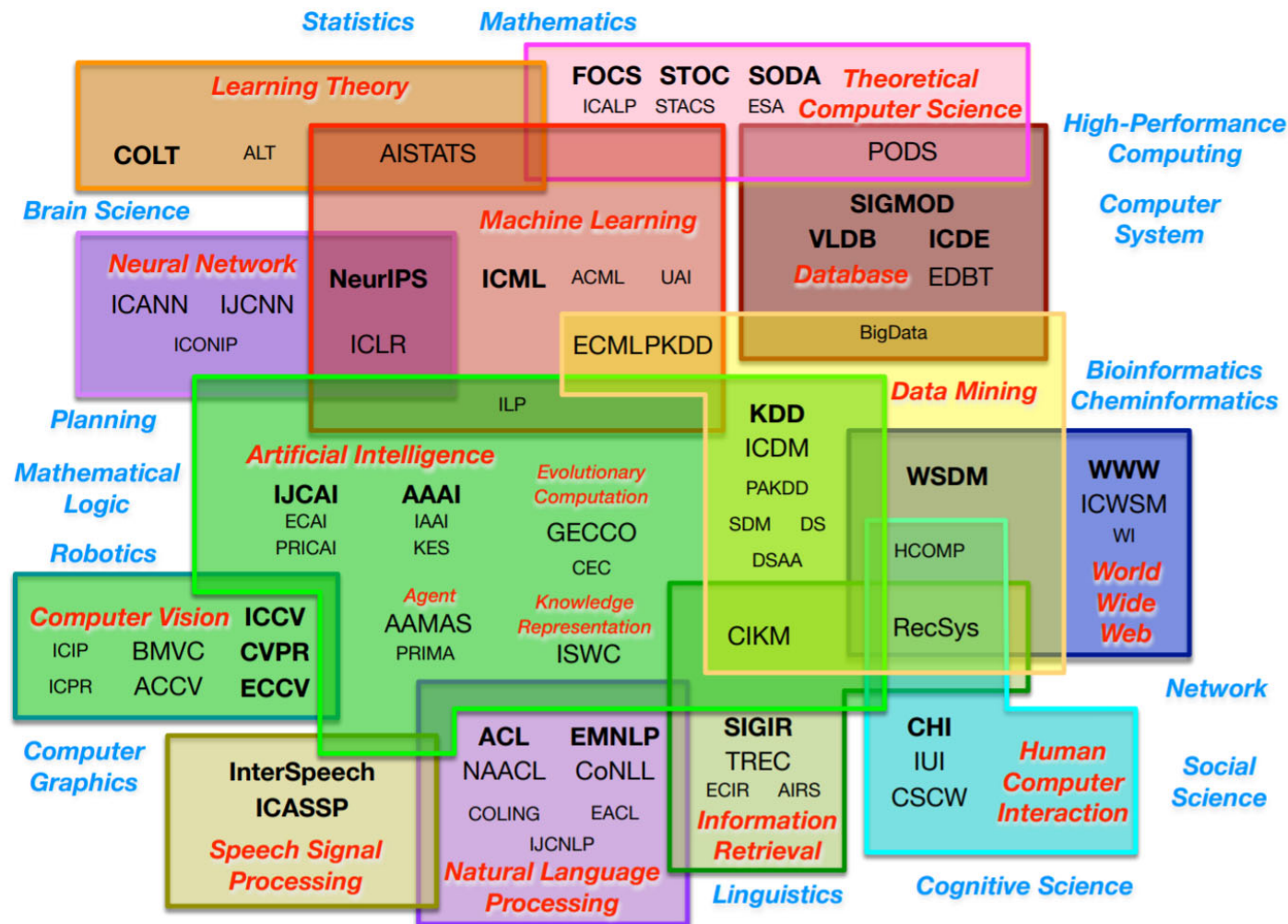
- 知能の要素技術を
更に高度化・統合化
- 人間のようなAI?



「人工知能」に関する国際会議

8

- ジャーナルでなく**国際会議**で研究成果を発表
- 基礎数学から実世界応用まで, 様々な学会が存在

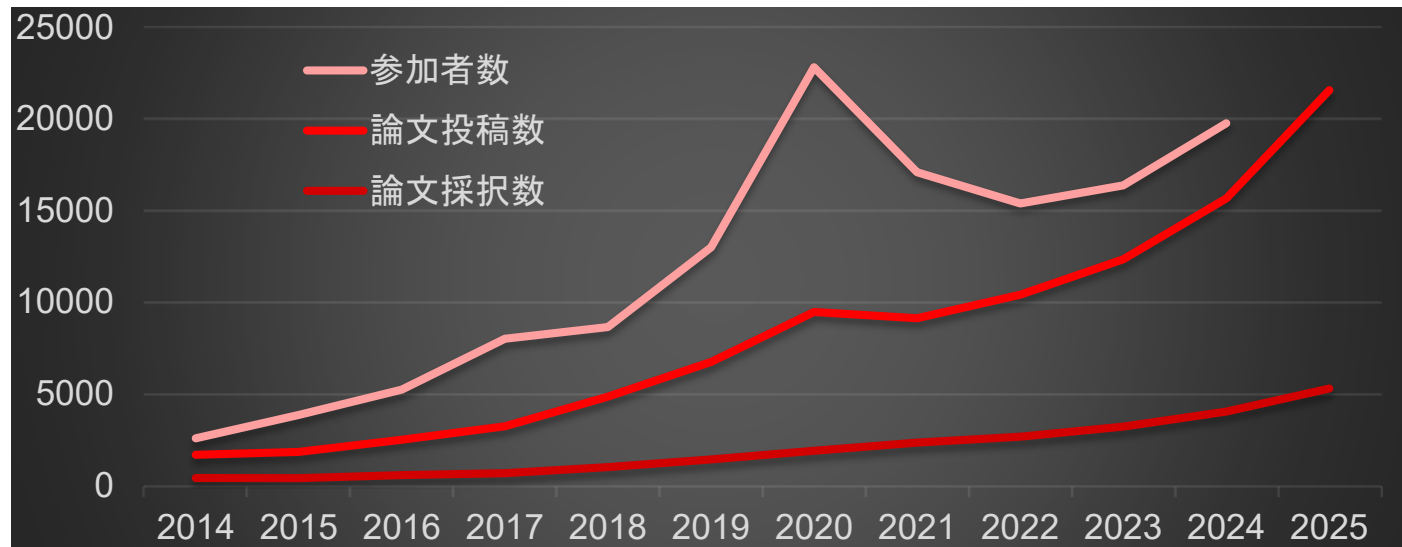


ML, DM, and AI Conference Map. Copyright © 2015 Toshiko Kamishima All Rights Reserved. Updated 2019-05-11

NeurIPSの動向



■ 参加者数，論文投稿数が激増：



https://media.neurips.cc/Conferences/NeurIPS2023/NeurIPS2023-Fact_Sheet.pdf

■ 企業のスポンサーも非常に活発：

- 2000年代前半：アメリカの大手IT企業
- 2000年代後半：世界中の大手IT企業
- 2010年代：非ITを含む様々な業種のベンチャー～大企業
- 2020年代：スポンサーを縮小へ？

■ 2015年ころ：

- 機械学習**技術そのもの**の議論が中心
- アルファ碁，自動運転車，会話ロボットなどが登場し，**技術のさらなる発展**への期待が高まる
- 研究，ビジネスとも，**北米**の企業，大学が支配的

■ 現在：

- 機械学習の技術開発競争が更に激化
- 機械学習の他の**科学研究への応用**
- 公平性などの**社会課題への取り組み**
- **米中**の企業の競争が激化
- マイノリティの支援など**多様性の重視**へ：
Women in ML, Black in AI, Queer in AI,
Global South in AI, Indigenous (先住民) in AI, ...

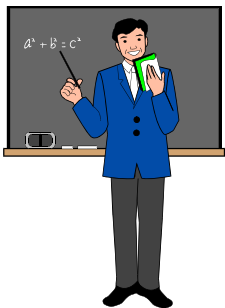
2025/1/10 日経新聞朝刊1面

「AIトップ論文著者数の上位機関」

NeurIPS2024, ICML2024, ICLR2024
の採択論文の著者の所属を調査

順位(20 年の順位)	機関名(国名)	採択論文 の著者数
1	(1) グーグル(米)	2174
2	(8) 精華大学(中)	1748
3	(2) スタンフォード大学(米)	1081
4	(3) マサチューセッツ工科大学(米)	1037
5	(7) カーネギーメロン大学(米)	1015
6	(17) 北京大学(中)	906
6	(89) 浙江大学(中)	906
8	(5) マイクロソフト(米)	851
9	(40) 上海交通大学(中)	810
10	(6) メタ(米)	782
⋮		⋮
64	(44) 理化学研究所(日)	188

- 米国の大学・企業と
中国の大学が上位を独占
- シンガポール, 英国, 韓国,
スイス, カナダなどの機関も
活躍
- 理研は20年と比べて2.1倍
増加したが, 順位は低下



講演の流れ

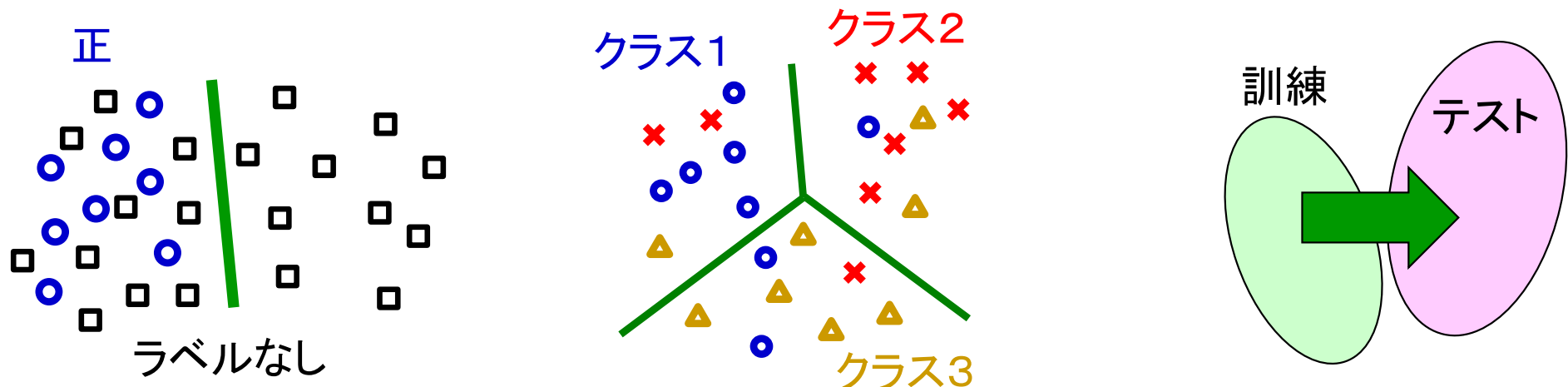
12

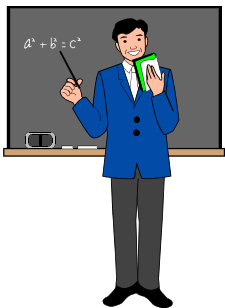
1. 機械学習の研究分野の紹介
2. 不完全な教師情報からの機械学習
 - A) 弱教師付き学習
 - B) ラベル雑音ロバスト学習
 - C) 転移学習
 - D) 最新の取り組み

最近の研究テーマ： 信頼できるロバストな機械学習

■ 様々な外乱の存在下で、
機械学習の信頼性を理論的に保証したい：

- 不十分な教師情報
- 雑音を含む教師情報
- バイアスを持つデータ





講演の流れ

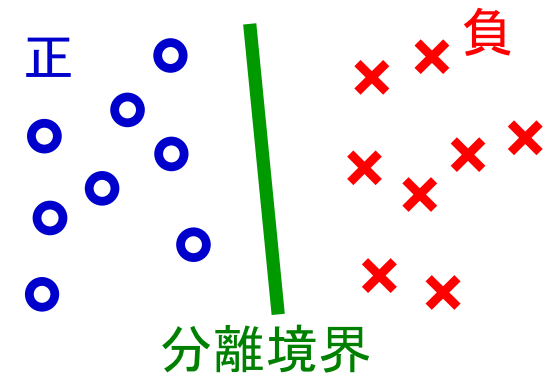
14

1. 機械学習の研究分野の紹介
2. 不完全な教師情報からの機械学習
 - A) 弱教師付き学習
 - B) ラベル雑音ロバスト学習
 - C) 転移学習
 - D) 最新の取り組み

■ 教師付き分類:

- 大量の良質な教師データを用いることにより, 人間と同等かそれ以上の予測性能を達成:
- 画像理解, 音声認識, 機械翻訳...

教師付き分類



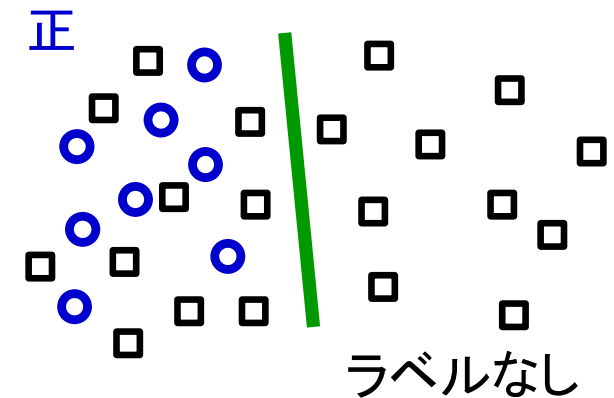
■ しかし, 応用分野によっては, 教師データを簡単に取れない:

- 医療, 自然災害, 材料, プライバシ...

■ 容易に集められる「弱い」教師情報を活用したい!

- 例: 正例とラベルなしデータからの分類

正ラベルなし分類

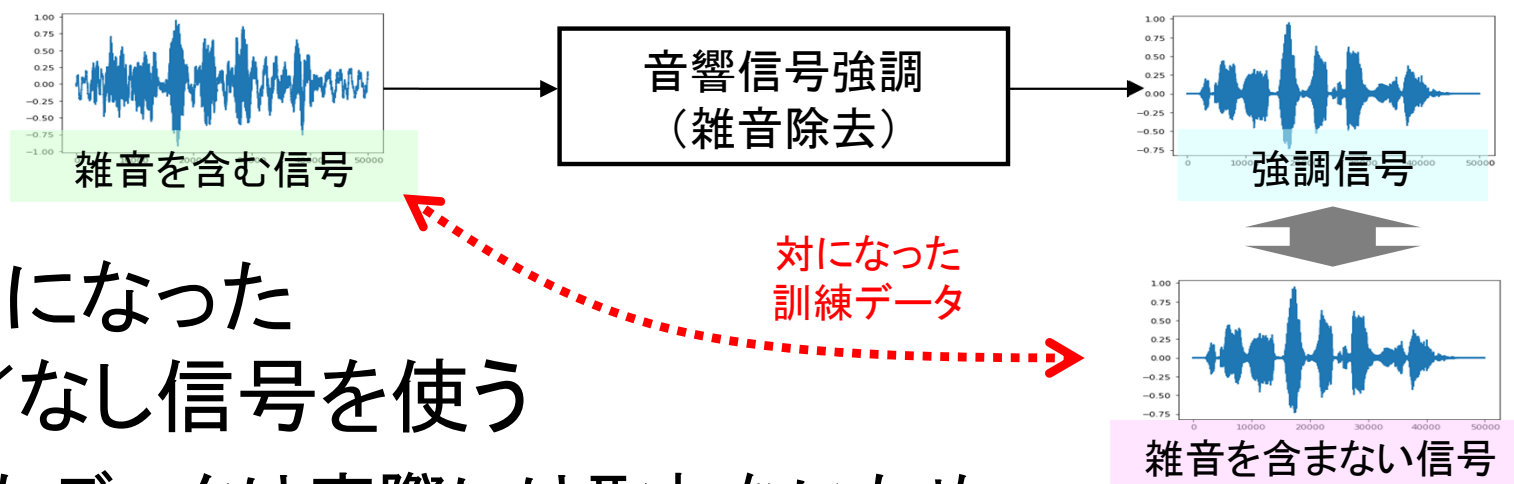


du Plessis+ (NeurIPS2014, ICML2015, MLJ2017),
Niu+ (NeurIPS2016), Kiryo+ (NeurIPS2017)

例: クリック予測

正ラベルなし分類の音響信号強調への応用 16

Ito & Sugiyama (ICASSP2023, Best Paper Award)



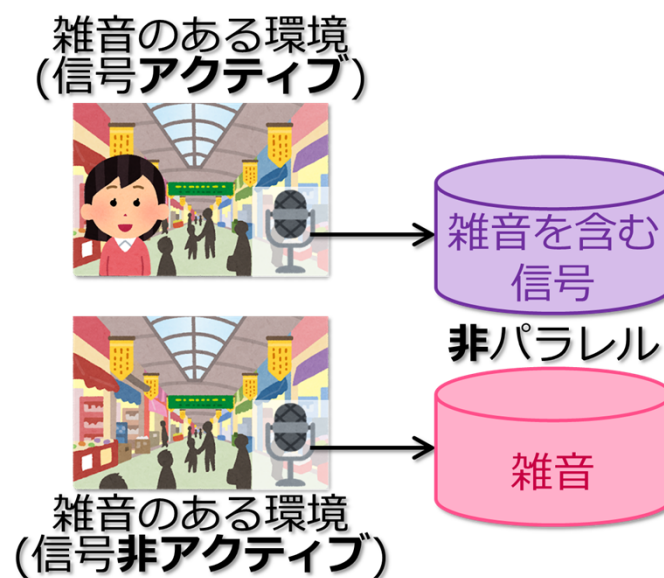
■ 従来法: 対になった雑音あり／なし信号を使う

- 対になったデータは実際には取れないため、人工データを使う → 実データに対してあまり汎化しない

■ 提案法: 対になっていない雑音あり信号と雑音から、正ラベルなし分類

対になっていない訓練データ	手法	SI-SNRi [dB]
	提案法	14.62 (0.20)
	MixIT ^{Wisdom+} (NeurIPS2020)	12.19 (4.50)
	教師付き学習	15.86 (1.28)

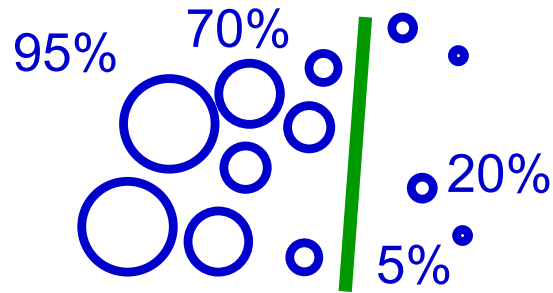
対になった訓練データ →



様々な弱教師付き分類

17

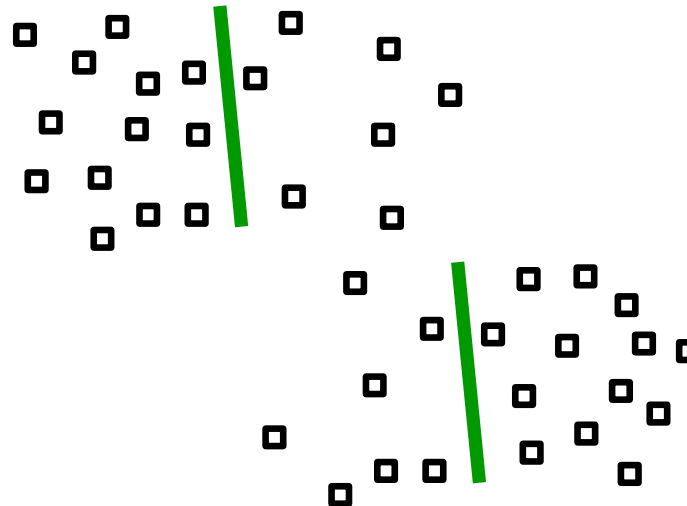
正信頼度学習



Ishida+ (NeurIPS2018), Shinoda+ (IJCAI2021)

例: 購買予測

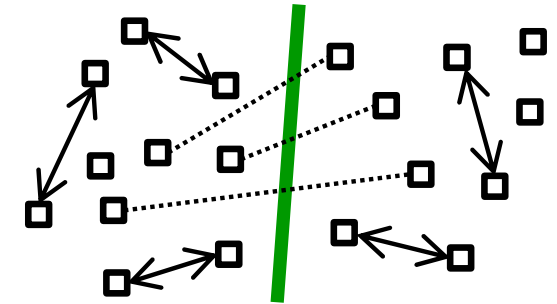
ラベルなしラベルなし分類



du Plessis+ (TAAI2013), Lu+ (ICLR2019, AISTATS2020), Charoenphakdee+ (ICML2019), Lei+ (ICML2021)

例: 異なる母集団からの学習

類似非類似ラベルなし分類



Bao+ (ICML2018), Shimada+ (NeCo2021), Dan+ (ECMLPKDD2021), Cao+ (ICML2021), Feng+ (ICML2021)

例: 機微情報予測

■ 多クラス分類へも拡張可能:

- 誤ったラベル, 曖昧なラベル...

Ishida+ (NeurIPS2017, ICML2019), Chou+ (ICML2020), Feng+ (ICML2020, NeurIPS2020), Lv+ (ICML2020), Cao+ (arXiv2021)

■ 任意の損失, 分類器, 最適化法, 正則化に適用可能!

■ さらなる発展:

- 統一的枠組み, 新しい問題設定, 新しい手法...

Chiang+ (TMLR2025), Chen+ (ICML2024), Lv+ (NeurIPS2024), Wang+ (NeurIPS2023, ICML2024, ICLR2025),

- 強化学習・模倣学習, 大規模事前学習モデル...

Cai+ (NeurIPS2023), Nishimori+ (RLC2025), Zhang+ (ICML2024), Li+ (MLJ2025), Lodkaew+ (TMLR2025)

「経験リスク最小化に基づく弱教師付き学習」

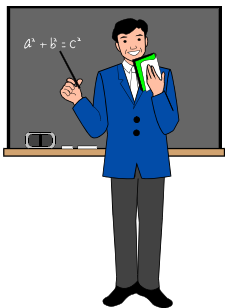
Sugiyama, Bao, Ishida, Lu, Sakai & Niu,
Machine Learning from Weak Supervision,
MIT Press, 320 pages 2022.

Machine Learning
from Weak Supervision

An Empirical
Risk Minimization Approach



Masashi Sugiyama, Han Bao,
Takashi Ishida, Nan Lu, Tomoya Sakai,
and Gang Niu



講演の流れ

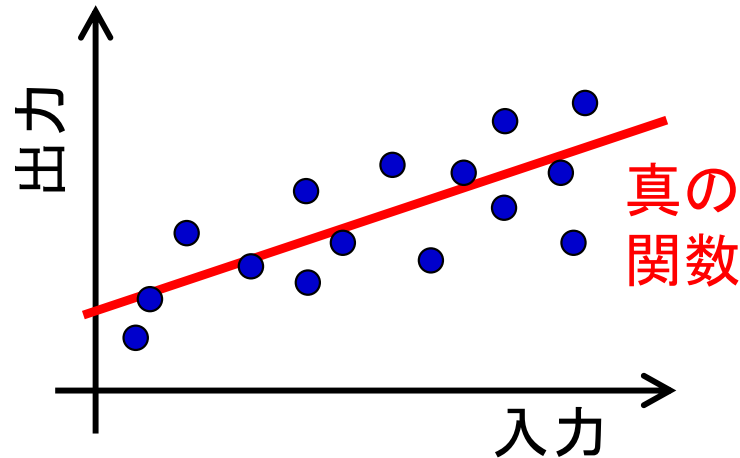
18

1. 機械学習の研究分野の紹介
2. 不完全な教師情報からの機械学習
 - A) 弱教師付き学習
 - B) ラベル雑音ロバスト学習
 - C) 転移学習
 - D) 最新の取り組み

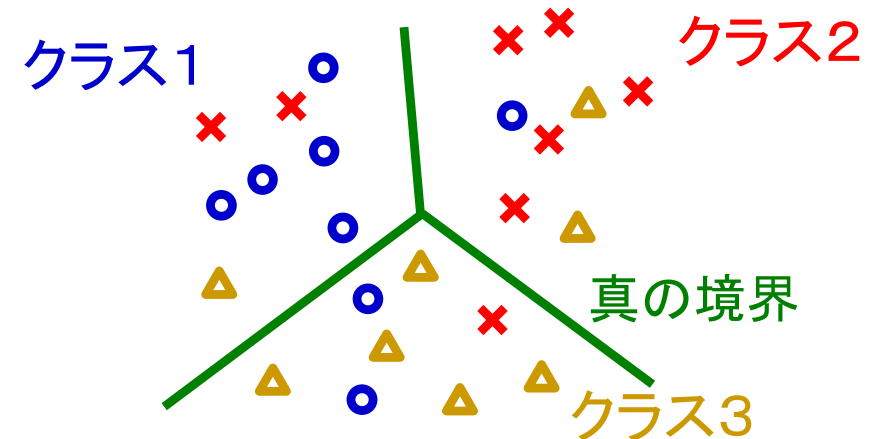
教師情報に含まれる雑音

19

回帰(加法雑音)



分類(ラベル反転)



- **回帰**: 単にデータを増やせばOK(一貫性がある)
- **分類**: データを増やしてもダメ(一貫性がない)
 - 明示的な**雑音除去機構**が必要!

ラベル雑音の補正

20

■ 雑音遷移行列:

- ラベル y が \bar{y} に反転する確率を表す行列
- これが分かれば雑音の影響を補正できる

Patrini+
(CVPR2017)

		1	0	0
y		0.1	0.8	0.1
		0.5	0.5	0
			\bar{y}	

■ 雑音を含むデータから雑音遷移行列を推定:

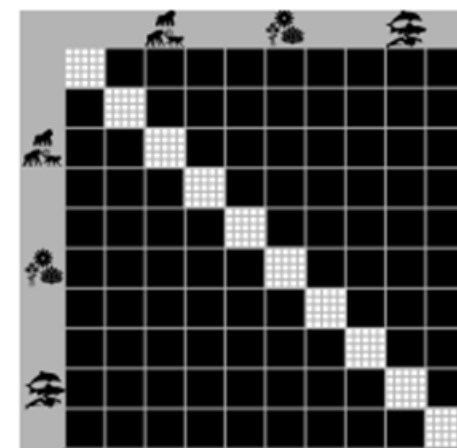
- ヒトの認知バイアスを活用
- 推定誤差の低減
- 分類器との同時推定
- 弱い仮定のもとでの一致推定

Han+ (NeurIPS2018)

Xia+ (NeurIPS2019)
Yao+ (NeurIPS2020)

Zhang+ (ICML2021)

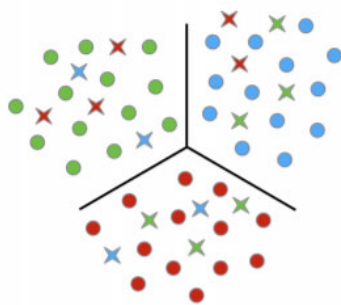
Li+ (ICML2021)



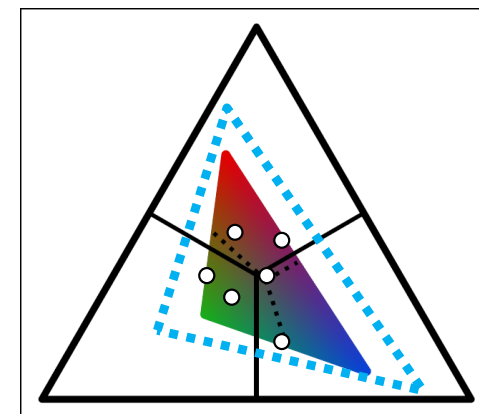
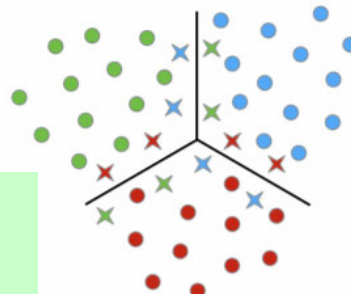
■ 入力依存雑音への拡張

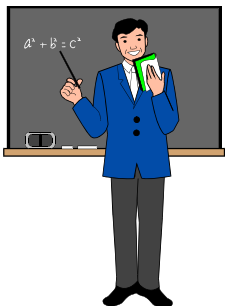
Xia+ (NeurIPS2020)
Berthon+ (ICML2021)
Cheng+ (CVPR2022)

入力
非依存



入力
依存





講演の流れ

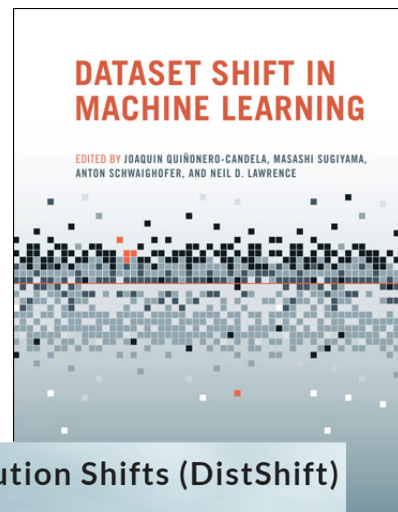
21

1. 機械学習の研究分野の紹介
2. 不完全な教師情報からの機械学習
 - A) 弱教師付き学習
 - B) ラベル雑音ロバスト学習
 - C) 転移学習
 - D) 最新の取り組み

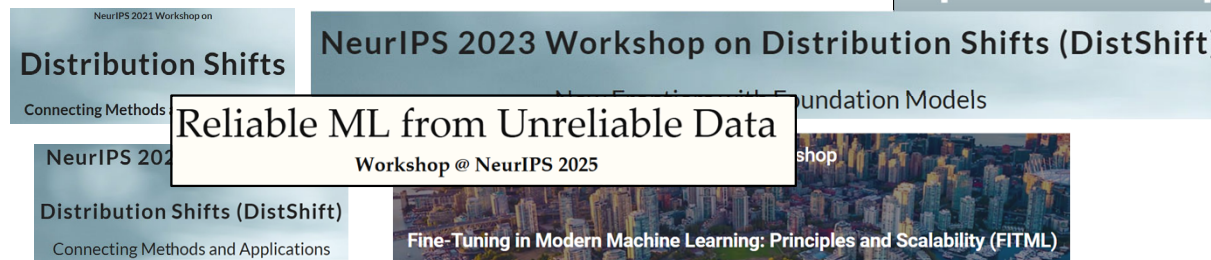
転移学習

■ 訓練データとテストデータの分布が異なると、標準的な機械学習法はうまくいかない：

- 環境変化, 標本選択バイアス(プライバシ)
- 訓練データをテストデータに適応させる



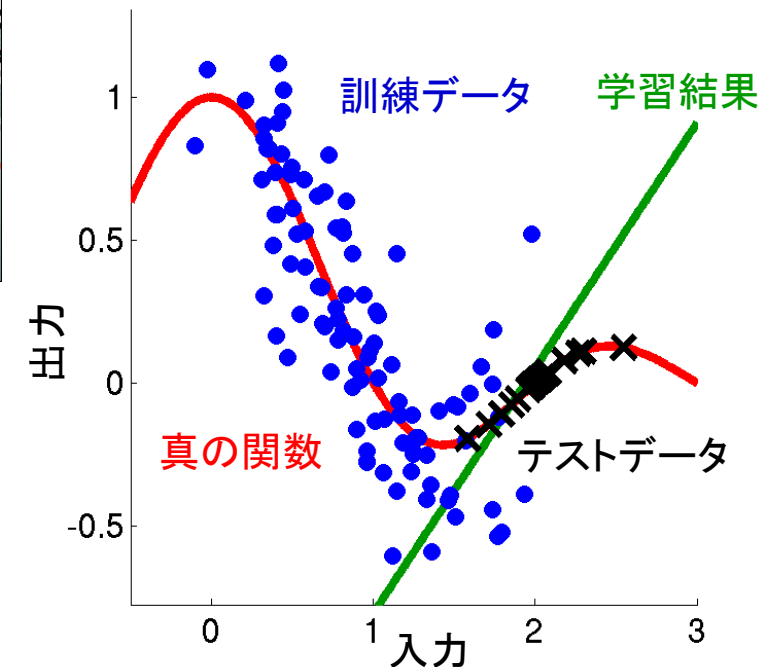
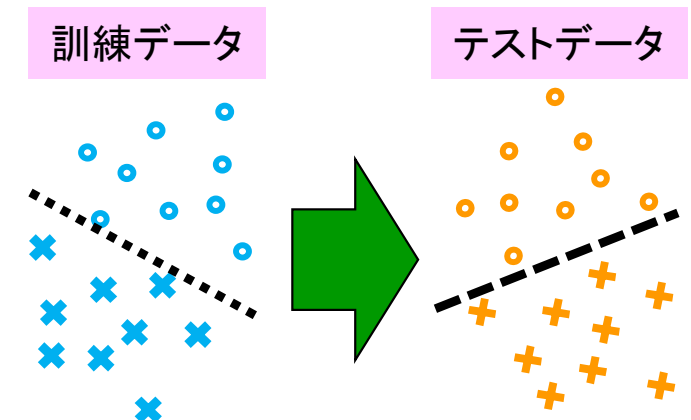
Quiñonero-Candela, Sugiyama
Schwaighofer & Lawrence (MIT Press 2009)



■ 典型的な設定：共変量シフト

Shimodaira
(JSPI2000)

- 入力分布だけが変化



重要度重み付け学習

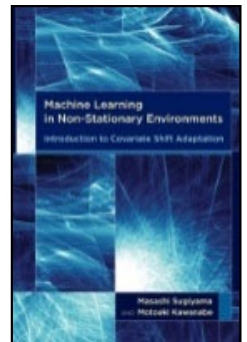
- 従来法: まず訓練データの重要度重みを推定し, 重み付け学習によって予測器を適応

$$\operatorname{argmin}_{f \in \mathcal{F}} \left[\sum_{i=1}^{n_{\text{tr}}} \frac{p_{\text{te}}(\mathbf{x}_i^{\text{tr}})}{p_{\text{tr}}(\mathbf{x}_i^{\text{tr}})} \ell(f(\mathbf{x}_i^{\text{tr}}), y_i^{\text{tr}}) \right]$$

\mathbf{x} : 入力

y : 出力

Sugiyama & Kawanabe
(MIT Press 2012)



- 新手法: 重みと予測器の同時学習

- 予測誤差の上界の同時最小化:

$$\min_{r, f} J_{\ell'}(r, f)$$

$$J_{\ell'}(r, f) \geq \frac{1}{2} R_{\ell}(f)^2$$

$$R_{\ell}(f) = \mathbb{E}_{p_{\text{te}}(\mathbf{x}, y)} [\ell(f(\mathbf{x}), y)]$$

$$\ell \leq 1, \ell' \geq \ell, r \geq 0$$

$$J_{\ell'}(r, f) = \mathbb{E}_{p_{\text{tr}}(\mathbf{x})} [(r(\mathbf{x}) - r^*(\mathbf{x}))^2] \leftarrow \text{最小二乗重要度推定}$$

$$+ (\mathbb{E}_{p_{\text{tr}}(\mathbf{x}, y)} [r(\mathbf{x}) \ell'(f(\mathbf{x}), y)])^2 \leftarrow \text{重要度重み付け学習}$$

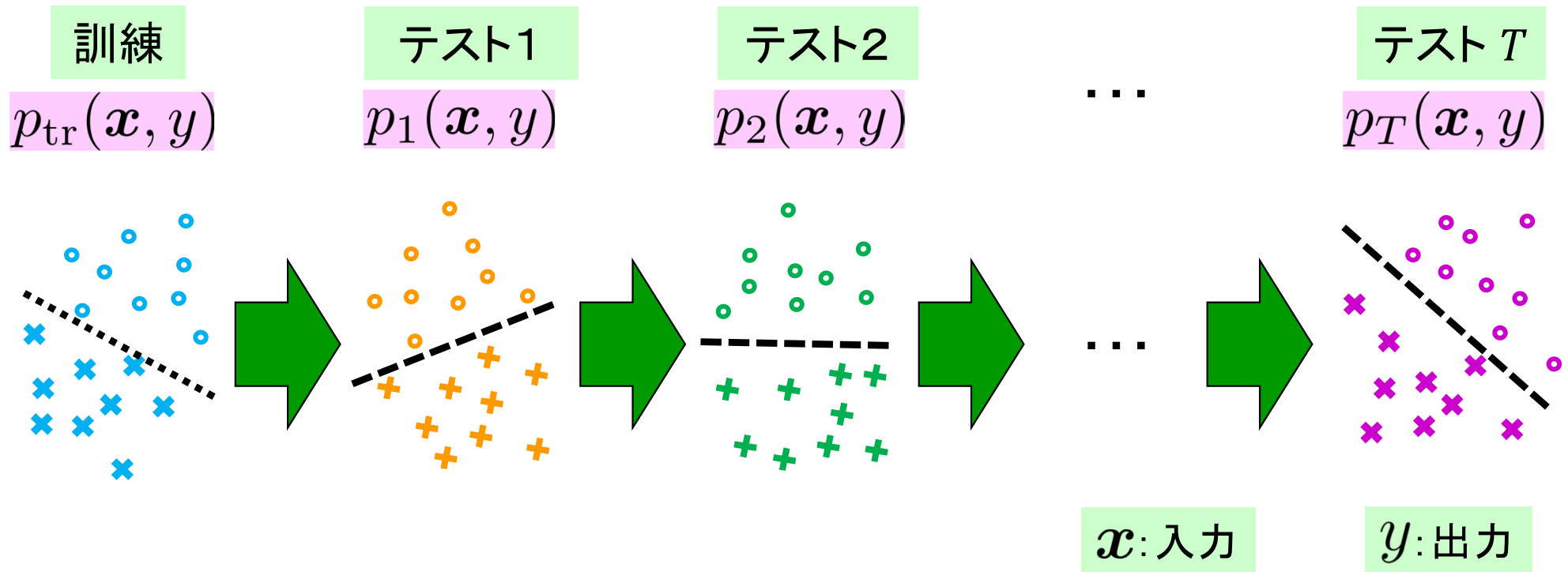
- 従来法は上界の二段階最小化に相当

- 収束性を理論保証:

$$\hat{f} = \operatorname{argmin}_{f \in \mathcal{F}} \min_r \hat{J}_{\ell'}(r, f)$$

$$R_{\ell}(\hat{f}) \leq \sqrt{2} \min_{f \in \mathcal{F}} R_{\ell'}(f) + \mathcal{O}_p(n_{\text{tr}}^{-1/4} + n_{\text{te}}^{-1/4})$$

Zhang+
(ACML2020 Best Paper
Award, SNCS2021)



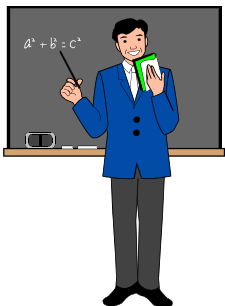
■ 連続クラス事前分布シフト: Bai+ (NeurIPS2022)

- クラスの比率 $p_t(y)$ だけが変化

■ 連続共変量シフト: Zhang+ (NeurIPS2023)

- 入力分布 $p_t(x)$ だけが変化

■ 動的リグレット最小化を理論保証: $\mathbb{E} \left[\sum_{t=1}^T R_t(\mathbf{w}_t) - \sum_{t=1}^T \min_{\mathbf{w} \in \mathcal{W}} R_t(\mathbf{w}) \right]$



講演の流れ

25

1. 機械学習の研究分野の紹介
2. 不完全な教師情報からの機械学習
 - A) 弱教師付き学習
 - B) ラベル雑音ロバスト学習
 - C) 転移学習
 - D) 最新の取り組み

同時分布シフト

- 転移学習の多くの手法では,
分布シフトの種類が決め打ち(例: 共変量シフト)

$$p_{\text{tr}}(\mathbf{x}) \neq p_{\text{te}}(\mathbf{x}) \quad p_{\text{tr}}(y|\mathbf{x}) = p_{\text{te}}(y|\mathbf{x})$$

- しかし, 分布シフトの種類の推定は難しい

- ラベル雑音も分布シフトの一種:

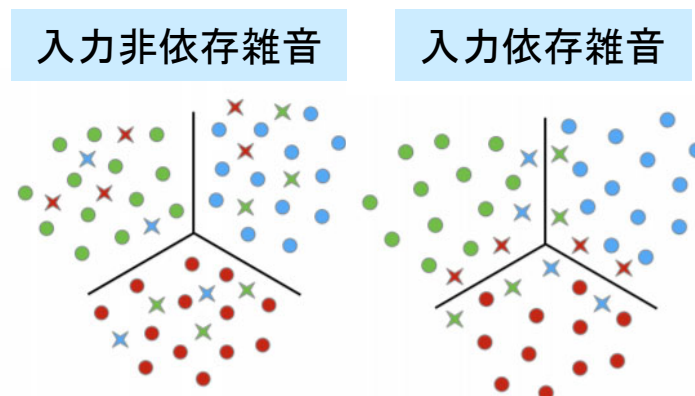
$$p_{\text{tr}}(\bar{y}|\mathbf{x}) = \sum_y \underbrace{p(\bar{y}|y, \mathbf{x})}_{\text{雑音遷移}} p_{\text{te}}(y|\mathbf{x})$$

\bar{y} : 雑音を含むラベル

- 入力非依存雑音に対しては良い理論と実用的な手法
- しかし, 入力依存雑音は難しい

- 同時分布シフトを考える:

$$p_{\text{tr}}(\mathbf{x}, y) \neq p_{\text{te}}(\mathbf{x}, y)$$



同時分布シフト:これまでとこれから 27

■ ラベル付きテストデータがあれば、なんとか解ける:

Fang+ (NeurIPS2020)

■ 与えられるデータ: 訓練とテストの入出力標本

$$\{(\mathbf{x}_i^{\text{tr}}, y_i^{\text{tr}})\}_{i=1}^{n_{\text{tr}}} \stackrel{\text{i.i.d.}}{\sim} p_{\text{tr}}(\mathbf{x}, y) \quad \{(\mathbf{x}_j^{\text{te}}, y_j^{\text{te}})\}_{j=1}^{n_{\text{te}}} \stackrel{\text{i.i.d.}}{\sim} p_{\text{te}}(\mathbf{x}, y)$$

■ 各ミニバッチ $\{(\bar{\mathbf{x}}_i^{\text{tr}}, \bar{y}_i^{\text{tr}})\}_{i=1}^{\bar{n}_{\text{tr}}}, \{(\bar{\mathbf{x}}_j^{\text{te}}, \bar{y}_j^{\text{te}})\}_{j=1}^{\bar{n}_{\text{te}}}$ に対して、重要度をカーネル平均適合で推定:

Huang, et al.
(NeurIPS2007)

$$\frac{1}{\bar{n}_{\text{tr}}} \sum_{i=1}^{\bar{n}_{\text{tr}}} r_i \ell(f(\bar{\mathbf{x}}_i^{\text{tr}}), \bar{y}_i^{\text{tr}}) \approx \frac{1}{\bar{n}_{\text{te}}} \sum_{j=1}^{\bar{n}_{\text{te}}} \ell(f(\bar{\mathbf{x}}_j^{\text{te}}), \bar{y}_j^{\text{te}})$$

■ ドメイン外データへの対応も可能:

Fang+ (NeurIPS2023)

■ ドメイン外への拡張:

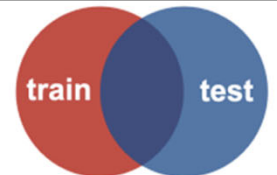
- 訓練ドメインの外では重要度が発散
- 外れ値検知を用いて、テストデータを訓練ドメイン内外に分割:

$$\{(\mathbf{x}_j^{\text{tein}}, y_j^{\text{tein}})\}_{j=1}^{n_{\text{tein}}}, \{(\mathbf{x}_j^{\text{teout}}, y_j^{\text{teout}})\}_{j=1}^{n_{\text{teout}}}$$

- 損失を個別に計算:

$$\frac{n_{\text{tein}}}{n_{\text{tr}} n_{\text{te}}} \sum_{i=1}^{n_{\text{tr}}} \frac{p_{\text{te}}(\mathbf{x}_i^{\text{tr}}, y_i^{\text{tr}})}{p_{\text{tr}}(\mathbf{x}_i^{\text{tr}}, y_i^{\text{tr}})} \ell(f(\mathbf{x}_i^{\text{tr}}), y_i^{\text{tr}}) + \frac{1}{n_{\text{te}}} \sum_{j=1}^{n_{\text{teout}}} \ell(f(\mathbf{x}_j^{\text{teout}}), y_j^{\text{teout}})$$

$$\frac{p_{\text{te}}(\mathbf{x}, y)}{p_{\text{tr}}(\mathbf{x}, y)}$$



■ これを弱教師付き学習・連続分布シフトの設定で解きたい

- 強い仮定無しで使える手法は、現場で最初に使うべき方法!

