

AI・犯罪等統計予測を 明日の社業に 如何に活かすか

西 颯人

一橋大学

ソーシャル・データサイエンス教育研究推進センター
特任助教

2022/12/13 オンラインJEASセミナー

1

自己紹介

- 西 颯人(ニシ ハヤト)
 - 特任助教
 - 一橋大学
ソーシャル・データサイエンス教育研究推進センター
 - 博士(工学)
- 出版物
 - 守山正(編著). 犯罪予測 —AIによる分析—. 成文堂.2022年
 - 西颯人、守山正. 2022. 5章:犯罪分析と統計学,
7章:AI活用による犯罪予測
 - 清水千弘(編著). 不動産テック. 朝倉書店. 2020年
 - 西颯人、清水千弘. 2020.
4章:不動産市場分析における統計・機械学習の利用

2

目次

- 警察のAI活用
- 統計とAI
 - どのように学習するか？－最尤推定－
 - カーブ・フィッティング
- 人間とAI
 - 機械学習の大枠
 - 何が違うか？
- 活用のスケール感

3

お話しすること

- 「AI」の作成を外注するのに
知っておいた方が良いこと
 - 発注側がある程度わかっているととてもスムーズ
- 「AI」や統計を勉強する前に知っておくと
助けになるかもしれないこと
- 巷で話題の「AI」の成果と、
犯罪分析・予測とで違うこと

4

警察のAI活用

5

犯罪対策への統計・AIの活用

- 京都府警
- 福岡県警
- 警察庁
- 神奈川県警
 - 産学官連携による人工知能を活用した犯罪・交通事故発生予測技法の調査研究
 - ↑講演者が調査研究補助者として参加
 - [犯罪や事故予測、AI活用報告書 県警、五輪前に運用も](#)
[| 社会 | カナロコ by 神奈川新聞 \(kanaloco.jp\)](#)

6

犯罪予測

- 事前予測と被害回避
 - あらかじめ予測できれば、犯罪被害を回避できる
 - 捜査にかかる時間と労力を削減できる
- 地域の問題解決
 - 環境犯罪学
 - SARA model
 - Scanning / Analysis / Response / Assessment
 - 地域で犯罪を誘発する要因を突き止めて、除去し解決していくことで、発生を抑止する

7

神奈川県警による犯罪予測

- 時空間メッシュ単位で犯罪の発生を予測
 - 100m×100m×3時間
- 対象罪種
 - 空き巣
 - 女性・子ども対象の性犯罪・前兆事案
 - 振り込め詐欺前兆電話
 - ひったくり
 - 自転車盗

8

プライバシー・権利保護

- フィクションでよくある「犯罪予測」
 - 「人」に対して犯罪者になるかを予測
 - その人のデータを持ってくる必要がある
 - そもそもまだ犯していない犯罪を裁けるのか？
- データの利用・実際の活用の双方で十分な配慮が必要

9

「地理的」犯罪予測

- 「地理的」犯罪予測
 - 「どこで」犯罪が起こるのかを予測する
 - 大山 智也, 雨宮 護, 島田 貴仁, 中谷 友樹, 地理的犯罪予測研究の潮流, GIS-理論と応用, 2017, 25 巻, 1 号, p. 33-43, 公開日 2019/12/31, Online ISSN 2185-5633, Print ISSN 1340-5381, <https://doi.org/10.5638/thagis.25.33>
 - 環境犯罪学との親和性がある

10

小括

- 海外はもちろん、国内の警察でも統計・AIの活用が試みられている
- 犯罪学的には、環境犯罪学が背景としてある
- 活用には十分な配慮が必要
 - 近年は犯罪予測への風当たりも強い面もある...

11

統計とAI

12

万引きの統計分析

どういう棚で万引き
が起こりやすいか？

- 藤本典志, 大山智也, and 雨宮護. "総合リユース店舗における万引き被害の空間的特徴 一般化線形混合モデルを用いた分析." *都市計画論文集* 53.3 (2018): 603-609.
- 藤本典志, 大山智也, and 雨宮護. "高ロス率商品に対するレジからの監視性確保による万引き抑止効果の検証 総合リユース店舗における店舗レイアウト変更のランダム化比較試験." *日本建築学会計画系論文集* 87.797 (2022): 1130-1140.

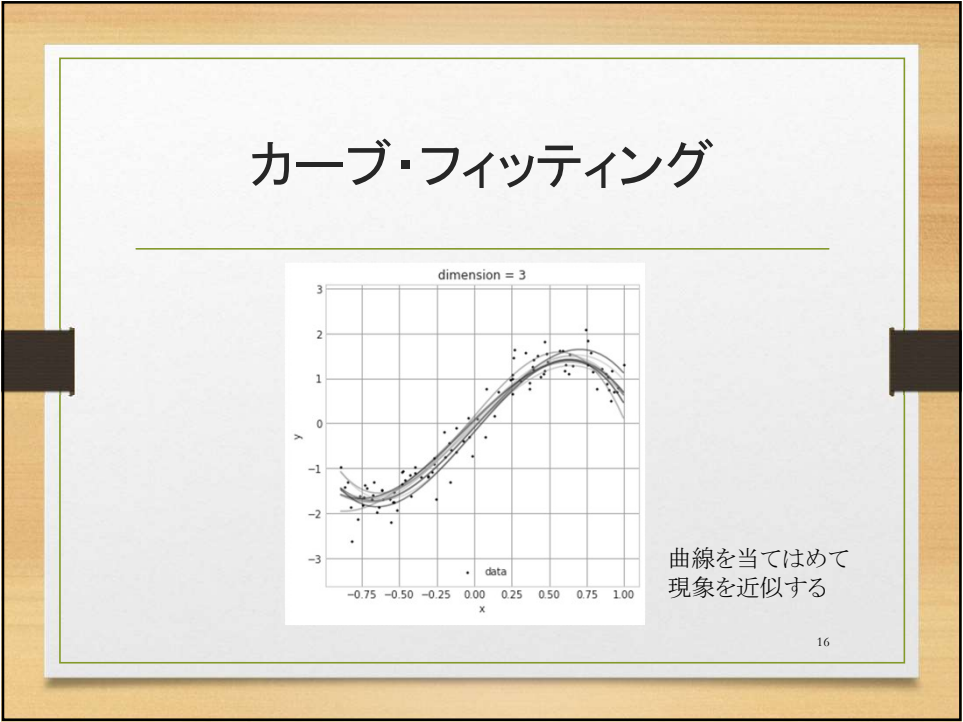
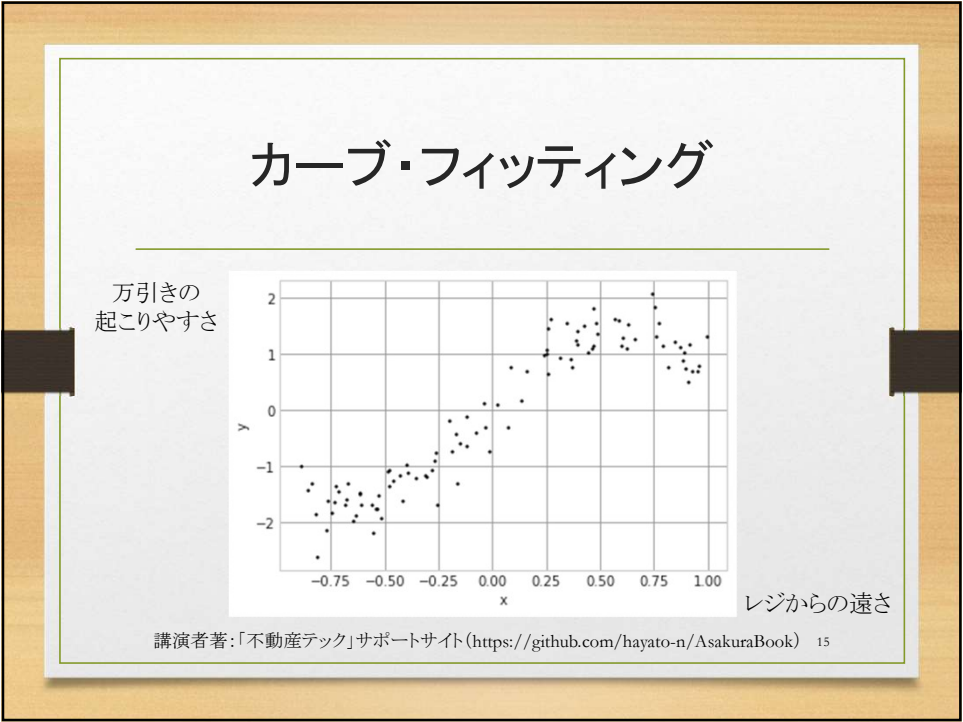
棚の配置を工夫したら
万引きは減るか？

13

統計モデルか、機械学習か

- 明確な違いはない
- なんか複雑そうなのが「機械学習」？
- どちらかという、統計は「現象の説明」に関心が、機械学習は「現象の予測」に関心がある
- よって、モデルの使い道に違いが出てくるかも...

14



どのように学習するか？

- 例:最尤推定
 - 使用するモデルに対して、適当なパラメータを設定する
 - そのパラメータを用いて、(使用するモデルから)所与のデータが得られる確率を計算する
 - その確率をパラメータ(とモデル)の「尤もらしさ」と考えて、確率が最大になるようにパラメータを調整する
- かなり汎用的で基本的な推定方法
 - 例:最小二乗法

17

何のために曲線を当てはめるか？

- 説明のため
 - 傾向をつかむため
 - 例:レジに近いほど万引き被害が少ないと言えるか？
 - 曲線が単純な方が傾向はつかみやすい(直線など)
- 予測のため
 - なんてかわかんけど、ここでは万引きが起こりやすいらしい！(お告げ)
 - 予測が当たるのが正義(傾向のつかみやすさは関係ない)

18

どう活用するか？

- 説明のため
 - どういう場所で起こりそうか傾向が分かるので、状況の改善につなげやすい
 - 予測の精度は犠牲になる場合も
- 予測のため
 - 予測に従い、「万引きが起こりそう」な場所を注視する
 - 予測の根拠を示そうという研究も盛ん

19

ご参考：検定

- 統計的に「関係があるか」言及するための仕組み
- 背理法的な論法
 - まず、「関係がない」と仮定する
 - 仮定の下で、観測されたデータが得られる確率を計算
 - 確率が小さい(5%以下など)なら、「関係がない」という仮定が間違っていたと考える
 - 確率が小さくない場合、「関係があるかわからない」

20

人間とAI

21

AI(人工知能)って何？



この写真の作成者 不明な作成者は CC BY-SA のライセンスを許諾されています

22

AI(人工知能)って何？

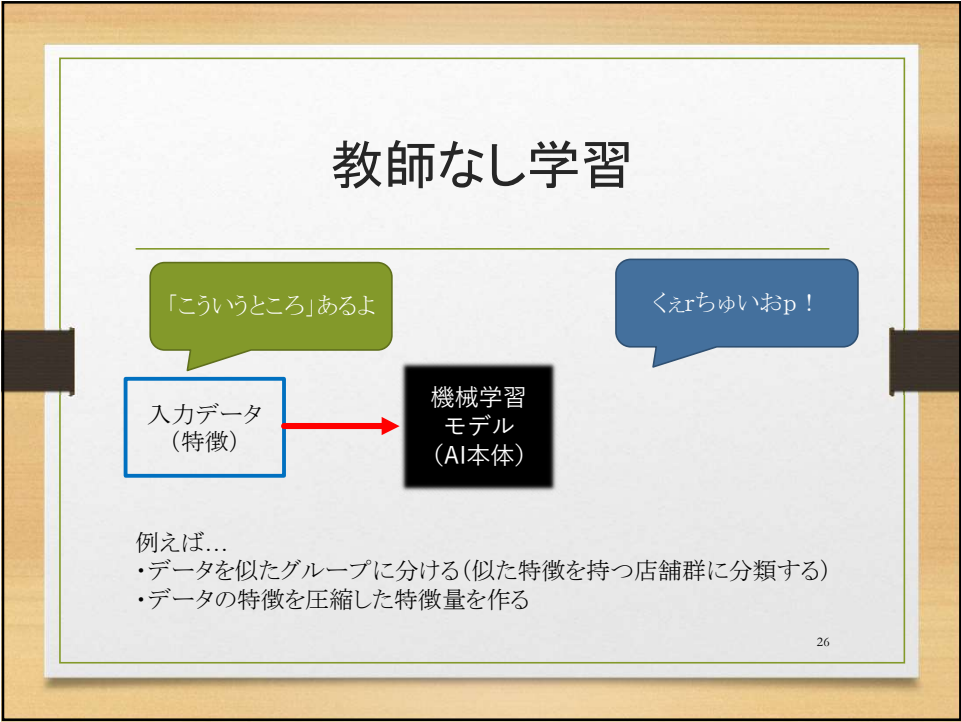
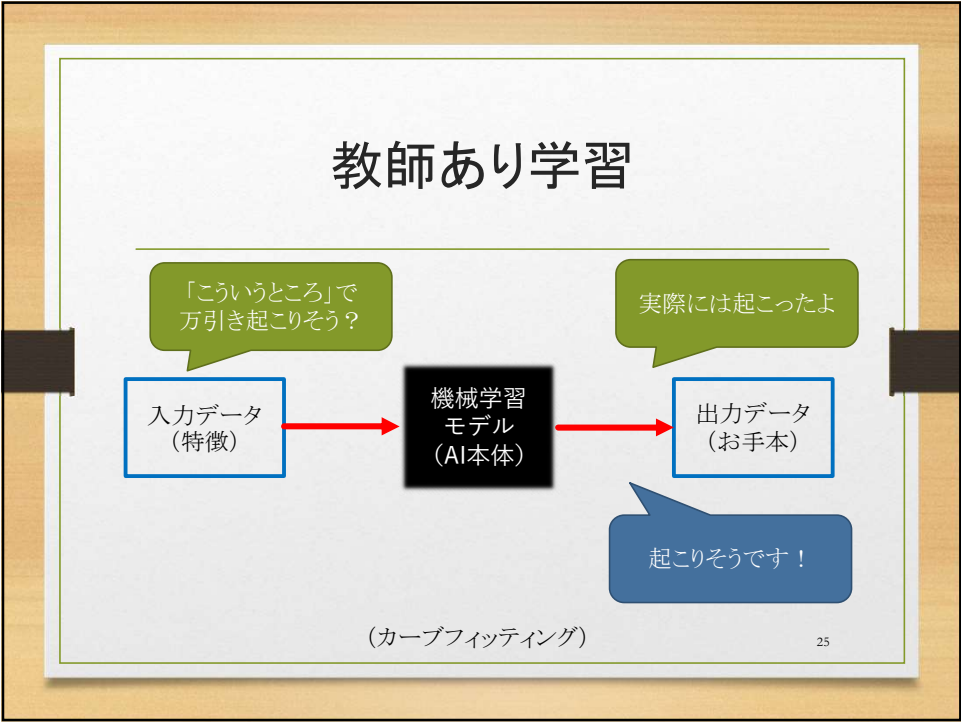
- 明確な定義はない
 - (自分で考えているっぽい感じがする機械？)
 - 商業的な理由から人工知能を名乗る場合も
- 機械学習？(私見)
 - 近年はニューラル・ネットワークなどの機械学習手法を使用したシステムをAIと呼称することが多い気がする
 - ニューラル・ネットワーク((人工)神経網)は人間の脳を模倣した構造なので、「人工知能」と名乗るのに親和性が高い気がする

23

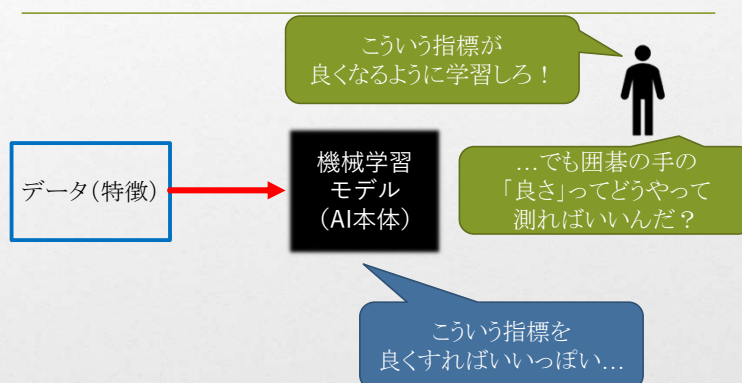
機械学習の枠組み

- 大きく3種類の枠組みがある
 - 実際にどのように「学習」するのは、ここではブラックボックスとします
 - 実用的にも、永久にブラックボックスになっているかもしれません

24



強化学習



強化学習

- Alpha Goが人間に勝ったらしい...
 - 強化学習の典型例
 - 囲碁の手の良し悪しの評価は困難
 - AI同士でとてつもない回数繰り返し対局する
 - 大量の棋譜と勝敗のデータが得られる
 - ここから「どの手が良かったか」を学習する
- 万引きは人間がやるので、
こういう学習の仕方はできない

28

ご参考:生成モデル

- お絵描きAIが話題...
- 大量のイラストをAIに見せて、似たようなイラストを生成できるように学習させる
- c.f.画像認識
 - 大量の画像をAIに見せて、そのラベル(写っているのが犬か猫か、など)を教えた通りに判別できるようにする
教師あり学習
- いずれも「画像」は大量に取得できるが、
泥棒はそんなにたくさんいない

29

現実的な方針

- 「学習」すべきことが多いほど、
データがたくさん必要
 - 人間が分かることは人間が教えてやる
- 基本的には「教師あり学習」を使う
- モデルの「良さ」は人間が決める
 - ⇒活用の結果どうなったらうれしいか、それをどうやって測るのかをあらかじめ決めておく

30

AIは賢く、空気を読まない

- 「万引きが起こる棚を全部予測してくれ！」
- AI
「全部の棚で万引きが起こるって予測すれば、全部予測したことになるな...」
- どんなモデルが「良い」のか評価するのは意外と難しい

31

人間への命令

レジから見えにくい棚
だと万引きされそう...

値段が高くて小さい
商品は盗まれそうだな...

このデータを使って万引き
が起こりそうか予測して？

棚の位置
陳列されている商品
etc...

経験に基づく推測と判断

32

AIへの命令

実際に万引きが
起こった棚と似た
棚を探すか...



このデータを使って万引き
が起こりそうか予測して？



棚の位置
陳列されている商品
etc...

推測・判断のもとになる人生経験がない
与えられた情報がすべて

**Garbage in,
Garbage out**

「西の朝食メニューは関係なさそうだ」という判断にもデータが必要
「与えられていないけど棚の高さも関係しそう」という判断ができない

c.f.ボードゲームは盤面にすべての情報が入っている

33

小括

- AIと人間は違う！
- AIを擬人化して考えるのはあまりよろしくない...
- AIに何をさせたいのかが大事
- AIの出力がどうなったらうれしいのか決めておく
- AIに入れる情報は人間があらかじめ精査するべき

34

活用のスケール感

35

活用の空間スケールを考える

- 活用場面は？
 - どの棚で万引きが起こりそうか？
 - どの店で万引きが起こりそうか？
 - どの地域で万引きが起こりそうか？
- 使うべきデータと活用方法が違う！
 - 序盤で挙げた統計分析例はマイクロ
 - 警察(神奈川県警)の「犯罪予測」はマクロ

マイクロ
↑↓
マクロ

36

どの棚で万引きが起こりそうか？

- 用意するデータ
 - 入力:
棚の位置、どういう棚か(高さ、おいてあるもの)、時点 etc.
 - 出力:
実際に盗まれたか？
- 活用方法
 - 統計:どんなふうに棚配置などを改めれば予防になるか
 - 機械学習:どの棚を監視すれば万引きを見つけられるか

37

どの店で万引きが起こりそうか？

- 用意するデータ
 - 入力:
店の特徴、立地、周辺施設、客層、etc.
 - 出力:
その店で万引きが起こるか or 何件起こるか
- 活用方法
 - 統計:どういう店で万引きに警戒が必要か、
近隣にどういう施設ができれば注意が必要か
 - 機械学習:どの店で警戒すれば万引きを見つけられるか

38

どの地域で万引きが起こりそうか？

- 用意するデータ
 - 入力:
地域特性、人口構成、アクセス性、施設立地、etc.
 - 出力:
その地域で万引きが起こるか or 何件起こるか
- 活用方法
 - 統計:どんな地域で万引きに警戒が必要か、
万引きを抑制するにはどんな介入が必要か
 - 機械学習:どの地域を警戒すれば万引きを見つけられるか

39

小括

- やりたい万引き対策によって、
空間スケールが変わってくる
- 空間スケールによって、
用意すべきデータが違ってくる
- やりたいことは何か、使えるデータは何か、を
あらかじめはっきりさせる必要がある

40

総括

- 現象(万引き発生)の説明か、予測か
 - 統計分析をするか、機械学習するか
- どのようなスケール感で「活かす」か
 - ミクロ？マクロ？
- AIにすべてをゆだねることはできない
 - たくさんのをやらせるには、たくさんのデータが必要
 - 人間がサポートしないとイケない
 - →社業を「補助」するような活かし方
 - 「AIを使わない」というのも1つの選択肢
 - 単純集計もバカにできない

41