

情報処理の応用B

第1回 ガイダンス

データサイエンス学部 成塚拓真

14:35開始

Takuma
Narizuka



講師

博士(理学)

成塚 拓真

なりづか たくま

専攻 (担当分野)

スポーツ統計科学、統計物理学、社会物理学

経歴：埼玉県鴻巣市出身，深谷市在住

- 2005～2008：早稲田大学本庄高等学院
- 2008～2017：早稲田大学先進理工学部物理学科（学部～博士）
- 2017～2021：中央大学理工学部物理学科（助教）
- 2021～現在：立正大学データサイエンス学部講師
駅伝部副部長

扱う内容

- 統計学の入門的な内容を網羅的に学ぶ
- 簡単なデータ解析ができるようになることを目指す
- 講義の構成
複数のテーマから成る
講義1回で1テーマ（講義ノートの1セクション）
- Pythonによるコーディングの例も示す
※プログラミングの詳細には立ち入らない

関係する講義

- 2年時必修科目「統計学I, II」
- 「情報処理の応用A」：レポートや評価基準は近づけるが内容は違う
- 「情報処理の基礎」：つながりはあまりない

講義資料

- 講義ノート，スライド，jupyter notebook (gogle colab)
⇒ ポータルサイト「オンライン授業」で適宜共有
- メインテキスト：
総務省政策統括官政策基準部編，「高校からの統計・データサイエンス活用～上級編～」，日本統計協会，2017
⇒ [PDFをDL可能](#)（購入希望者は来週までに成塚までメール）
- 参考文献
竹村彰通・姫野哲人・高田聖治編，データサイエンス入門，学術図書出版社，2019.
東京大学教養学部統計学教室編，統計学入門，東京大学出版会，1991.
薩摩順吉，確率・統計，岩波書店，2019.
栗原伸一，入門統計学（第2版），オーム社，2021.

パターンI（対面の場合）

- 基本的には板書
- 適宜スライドを用いる
- Pythonが動くPCが手元にあると良い

パターンII（オンラインの場合）

- Zoomによる同時双方向授業
- 基本的にはスライド（手書きになるかも）

講義情報の共有方法

- ポータルサイトの「オンライン授業」を用いる
- 対面の場合も同様

レポート（60%）

- 2回くらい？
- （予定）講義で扱った実例についてレポートにまとめる
- （予定）添削して返却し，再提出してもらう
⇒ 文書を書く訓練を兼ねる

授業への取り組み姿勢（40%）

- 授業への出席で評価

試験

- 行わない

レポートの書き方

7

- 鉄則：“当事者でなくても分かるように書く”
- エントリーシートなどと同じ気持ちで書く
- Word, LATEX, などを使って書いても良いが、書式は統一する
(フォントや文字サイズを統一, 見出しをつける, 数式エディタを使う)
(ソフトを使いこなせないなら手書きのほうが良い)
- レポートの本文はフォーマルな文章で過不足なく記述する
(省略記号やメモ書き, パワポのスライドのような記述は相応しくない)
- 結果を表やグラフにまとめる場合は本文でも詳細を説明する
- 初めにレポートのタイトル, 学籍番号・氏名を書く
- ファイル形式はPDF
- レポートは1つのファイルにまとめ, ファイル名は“学籍番号_氏名.pdf”などとする

導入：データサイエンスの役割

参考文献

竹村彰通・姫野哲人・高田聖治編,
「データサイエンス入門」, 学術図書出版社, 2019.

ビッグデータ

- 情報通信技術や計測技術の発達で取得可能になった多種多様なデータ
- ネットワーク上に蓄積
- 様々な機器により（知らないうちに）収集されている

例：スマートフォン

- アップル社が2007年に発売したiPhoneによって確立
- インターネットに常時接続
- 携帯電話＋デジカメ＋音楽プレイヤー＋...
- 発売10年で個人所有率7割
- 30年前のスーパーコンピューターに匹敵

例：スマートフォン

- スマートフォンの普及 ← 無線通信の速度, 容量の増加
⇒ 地下鉄の中でも常に「圏内」
- 人々の行動を通じた様々なデータの取得
SNSを通じたメッセージの交換
ブラウザを用いた情報収集
電子マネーによる買い物 ⇒ 購買履歴の蓄積
SUICAなどで電車に乗る ⇒ 移動履歴の蓄積
- 取得されたデータの活用
インターネット上のサーバに蓄積
商品やサービスのトレンド分析
⇒ 年齢や性別による関心や消費行動の違いを分析
⇒ 新商品の開発への応用

例：人工衛星による観測

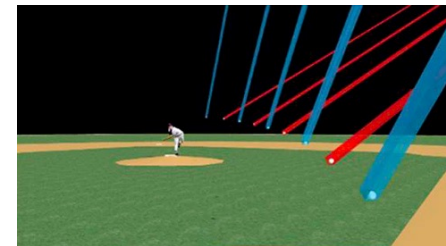
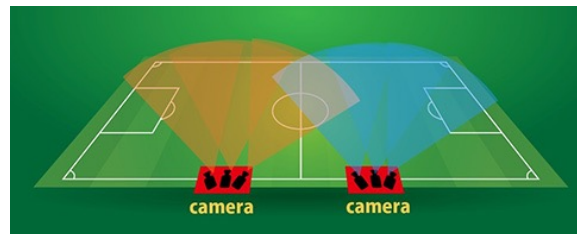
- 日本の気象衛星「ひまわり」
ひまわり初号（1977年）→ ひまわり9号（2016年）
- 観測性能の向上
観測頻度：3時間ごと → 2.5分ごと
分解能：1.25km → 0.5km
- 人工衛星を用いた位置測定
カーナビやスマートフォンの位置情報に不可欠
- GPS衛星「みちびき」
2017年に4号機の打ち上げ
4機体制 ⇒ 24時間いつでもデータを利用可
誤差数cm ⇒ 無人トラクターによる種まきや農薬散布への応用

例：センサー

- 今後重要性が増すと期待される
- スマートフォン
輝度センサー（明るさ調整）
モーションセンサー（画面の回転）
地磁気センサー（利用者の方向検出）
- 自動運転車
カメラやレーザー光を使った[LiDAR](#)（ライダー）
レーザー光が物体に当たって跳ね返るまでの時間で距離や方向を測定
- IoT（Internet of Things）
「モノ」をインターネットに接続する技術
IoT技術を生産現場に応用（＝スマート工場）⇒ [第4次産業革命](#)
[ソサエティー5.0](#)：サイバー空間とフィジカル空間の融合

例：スポーツ

- カメラやセンサーによるデータの自動取得
サッカー・バスケ：トラッキングデータ（2010年頃～）
野球：1球データ，PITHf/x
- スポーツ統計科学の発展
個人競技 ⇒ 対戦型スポーツ
統計科学，データサイエンスに基づく新たな分野



データスタジアム株式会社HPより転載 (<https://www.datastadium.co.jp/service/tracking.html>)

あらゆるデータを処理・分析して有用な情報（価値）を引き出す
＝ データサイエンス

資源としてのデータ

- データは新たな経済的資源 = 「21世紀の石油」
⇒ データを保有するものが有利
- GAFA（グーグル，アップル，フェイスブック，アマゾン）
ネット上の大企業
世界中で億単位のユーザを囲い込み
個人のデータを独自に収集・分析し，新たなサービスを展開
- 中国
政府の政策により，国内のインターネット事業者を保護
BAT（バイドゥ，アリババ，テンセント）
- データ駆動社会
資源としてのデータの利用がイノベーションをもたらしている社会

資源としてのデータ

- プラットフォーマー

GAFAやBATなど、サービスやシステムを提供する事業者

フェイスブック：SNSのプラットフォーマー

- ネットワーク効果

1つのサービスに加入している人が多いほど運営しやすい

⇒ 一旦ユーザが集まったサービスにはさらに多くのユーザが集まる

- 負の側面

個人情報の流出など

フェイスブックから8700万人の個人情報が流出

⇒ データ分析会社に渡り、2016年の米大統領選で悪用された？

データサイエンティスト

- データを処理・分析する技術の必要性
データを資源として貯めているだけでは価値を生まない
- 日本の現状
データを外国企業に取られ、活用されている
データを加工・分析する技術、人材の欠如
- データサイエンティスト
データを処理・分析し、データから価値を引き出せる専門的人材
- データサイエンティストに求められる素養
情報学、コンピュータ科学＋数学＋統計学（理系的）
応用分野は文系的
⇒ データサイエンスは文理融合

21世紀のそろばん

- 縦串の手法
固有の専門領域で使われる手法（スポーツ科学，観光学，法学）
- 横串の手法
分野を問わず必要とされる手法（数学，データサイエンス）
最近のインターネット関連のイノベーションに貢献

AI（人工知能）

- 深層学習（Deep Learning）の発展に伴い注目
- ビッグデータの存在と表裏一体
- 政府による「[AI戦略2019](#)」
数理・データサイエンス・AI教育の全国展開

- データサイエンティスト ⇒ 魅力的な職業
グーグルのチーフエコノミスト，上級副社長の発言
- アメリカの例
統計学・生物統計学の学士・修士号の授与数が2009~2019で大幅増
多くの大学に統計学科が存在 ⇒ データサイエンス教育にシフト
- 中国の例
300以上の大学に統計学部・学科が存在
巨大インターネット企業が多いデータサイエンティストを採用
- 日本の例
2017年に滋賀大にデータサイエンス学部開設（初の統計学専攻学部）
⇒ 毎年データサイエンス系の学部が新設（立正も）
多くの企業でデータサイエンティストの需要が激増

- 次回から（たぶん）対面 ⇒ 教室はA202
- 教科書購入希望者は成塚までメール
- 本日のスライド，講義ノートは後日共有
- 詳細はポータルサイトのオンライン授業