

データサイエンスやDXとはなんだろうか

木村巖（学術研究部理学系）

2022年3月24日（木）

目次

- ▶ データサイエンスとはなんだろうか
- ▶ DX とはなんだろうか

データサイエンスとは何だろうか

データサイエンスとは何だろうか

- ▶ 多くの大学で「データサイエンス」が必修化されている
- ▶ 本学でも
 - ▶ 教育・学生支援機構にデータサイエンス推進センターが設置
 - ▶ 教養教育科目「情報処理」でデータサイエンスの基礎を扱う
 - ▶ 各学部で「数理・データサイエンス・AI教育プログラム」が開設されている
 - ▶ 文部科学省「数理・データサイエンス・AI教育プログラム認定制度（リテラシーレベル）」認定されたプログラム

しかし、「データサイエンス」とは実際にはなにか？

データサイエンスとは

- ▶ 統一的な見解は.....?
 - ▶ 少なくとも「科学」ではあるはず
- ▶ 確率論と統計学
 - ▶ 数学の一分野として
- ▶ 大量のデータを取り扱う技術
 - ▶ コンピュータで大規模データを処理する技術

なぜいまデータサイエンスがこんなに流行っているのか？

なぜいま「データサイエンス」なのか

- ▶ データを収集しやすくなった
 - ▶ 以前は、政府・官公庁・企業の統計、POS システム、交通事業者、インフラ事業者など
 - ▶ 現在は、ほとんどの人がスマホを持つようになった
 - ▶ 写真、位置情報、インターネットへのアクセス状況など
 - ▶ さらに、「IoT」の時代へ
- ▶ データを処理しやすくなった
 - ▶ コンピュータの計算能力の拡大
 - ▶ ビッグデータを処理する技術の確立と普及
 - ▶ GPU (General Purpose processing Unit)
 - ▶ 使いやすいプログラミング言語とライブラリ
- ▶ 「データサイエンティスト」といった職業も生まれた
 - ▶ 高校までの数学のカリキュラムにも該当する単元がある

統計学

- ▶ 例：高校の課程に含まれる「統計」
- ▶ 「数学I」のカリキュラムに「データの分析」の単元がある
 - ▶ データの整理と分析
 - ▶ 代表値，箱ひげ図，分散と標準偏差
 - ▶ データの相関
 - ▶ 相関関係

統計のデモ

- ▶ Excel でやる例
 - ▶ 「分析ツール」アドインを有効にする
 - ▶ データの整理と分析
 - ▶ 代表値, 箱ひげ図, 分散と標準偏差
 - ▶ データの相関
 - ▶ 相関関係
- ▶ データは Kaggle で提供されている「**Students Performance in Exams Marks secured by the students in various subjects**」より.
 - ▶ **Kaggle** は, データ解析のためのプラットフォームサービス. Jupyter notebook 上で, Python の様々なライブラリを使ってデータ解析, 統計モデリングができる環境を提供している.

確率論

- ▶ 例：高校の課程に含まれる「確率・統計」
- ▶ 「数学 B」のカリキュラムに「確率分布と統計的な推測」の単元がある
 - ▶ 確率分布
 - ▶ 確率変数と確率分布，確率変数の期待値と分散
 - ▶ 確率変数の和と積
 - ▶ 二項分布
 - ▶ 統計的な推測
 - ▶ 母集団と標本
 - ▶ 標本平均の分布
 - ▶ 推定

確率のデモ

- ▶ Excel でやる例
 - ▶ 二項分布 $B(n, p)$, $n > 0$, $0 \leq p \leq 1$ が n が大きくなるにつれ正規分布に近づく例.

Python, pandas, matplotlib

- ▶ Excel ではなく，Python や関連ツールがいろいろあるみたいですが？
- ▶ Python, pandas, matplotlib をざっと紹介

Python

- ▶ Guido van Rossum が 1990 年台に開発を始めたプログラミング言語
 - ▶ 学習のしやすさ
 - ▶ 誰が書いてもほぼ同じになる
 - ▶ プログラムの構造を「タブ」(字下げ) で表現
 - ▶ 多くのウェブサービスで採用され人気が出た
 - ▶ Youtube, Instagram, Netflix, Spotify, etc.



Zscout370 - 投稿者自身による作品, パブリック・

pandas

pandas

- ▶ Python で様々なデータを読み込んで、加工して、保存したりグラフを描画したりするライブラリ。
- ▶ デファクトスタンダードとして広く使われている

matplotlib

matplotlib

- ▶ Python で様々なグラフを描くためのライブラリ
- ▶ こちらもデファクトスタンダード
- ▶ その他に, **seaborn** もあわせて広く使われる.
 - ▶ **seaborn** は **matplotlib** 上に構築された, データビジュアライゼーションのためのライブラリ.

Python などを使う理由は？

- ▶ Excel で問題なく仕事ができるならそれで OK
- ▶ Python などを使うと、データ処理の過程の「再現性」を維持するのが容易
 - ▶ Excel だと、データの加工の手続きを記録するのが手間
 - ▶ 例：csv データのある列とある列を取り出し、無効なデータを含む行を削除し、云々
 - ▶ Python や pandas を使うと、「データの加工の手続き」がプログラムそのもの
- ▶ データ処理の過程の「再現性」は非常に重要
 - ▶ 10 年間の保存が求められる

Jupyter notebook

Jupyter notebook

- ▶ 昔ながらのやり方：
 - ▶ プログラムファイルをエディタを使って書く
 - ▶ 黒い画面（端末ソフト）で実行する
 - ▶ グラフはグラフ描画用のソフトで描画し、表示する
- ▶ 今どきのやり方：
 - ▶ Jupyter notebook という、ウェブブラウザを用いたノートブックを用いる
 - ▶ **Jupyter**
 - ▶ もとは、IPython という名前で Python 専用が開発されていた
 - ▶ 他の多くの言語をサポートする仕組みが用意され、特定のプログラミング言語用ではなく汎用のノートブックソフトになった
 - ▶ 最近は Jupyter の次世代版 JupyterLab も普及し始めている。

環境構築が難しい

- ▶ Python, pandas, matplotlib, Jupyter, etc...
- ▶ こういったものを使えるようにする（環境構築する）のが難しい
- ▶ **Anaconda** はそういった一切合切を面倒見てくれる
 - ▶ インストーラ兼パッケージマネージャ
 - ▶ ただし，商用利用に関してはライセンスを確認すること
- ▶ **Google Colab**
 - ▶ Jupyter notebook を使える環境を無償提供
 - ▶ 有償環境もあり（強力な環境を長時間使える）
- ▶ **Kaggle** も同様
- ▶ **Amazon Sagemaker Studio Lab** も Preview 中.

Google Colab での実行例

- ▶ 学生のデータ（上述の Kaggle で提供されているもの）を Google Colab で調べてみる。

AI と機械学習

- ▶ AI と機械学習の違い
 - ▶ AI.....Artificial Intelligence, 人工的な知性の総称
 - ▶ 機械学習.....Machine Learning, 学習アルゴリズムとその実装の総称
- ▶ 機械学習の例
 - ▶ 回帰
 - ▶ 決定木
 - ▶ ディープラーニング
- ▶ **Scikit Learn, PyTorch, TensorFlow** といった著名なフレームワークを使うと簡単に実行できる

ここまでの参考文献

- ▶ 俣野博，河野俊丈編「数学Ⅰ」，東京書籍
- ▶ 中山浩太郎監修「東京大学のデータサイエンティスト育成講座」，マイナビ

DX とは何だろうか

DX とは何だろうか

▶ デジタルトランスフォーメーションを推進するためのガイドライン (DX 推進ガイドライン) Ver. 1.0

本ガイドラインでは、DX の定義は次のとおりとする。「企業がビジネス環境の激しい変化に対応し、データとデジタル技術を活用して、顧客や社会のニーズを基に、製品やサービス、ビジネスモデルを変革するとともに、業務そのものや、組織、プロセス、企業文化・風土を変革し、競争上の優位性を確立すること。」

大学における DX をどのように理解するか？

「企業がビジネス環境の激しい変化に対応し、データとデジタル技術を活用して、顧客や社会のニーズを基に、製品やサービス、ビジネスモデルを変革するとともに、業務そのものや、組織、プロセス、企業文化・風土を変革し、競争上の優位性を確立すること。」

- ▶ 営利企業ではない大学で、DX をどのように理解するか？
 - ▶ 「顧客」≡ 学生・保護者（？）
 - ▶ 「サービス」.....大学はサービス業か？
- ▶ 「業務そのものや、組織、プロセス、企業文化・風土を変革し競争上の優位性」
 - ▶ 我々は誰と競争しているのか？
 - ▶ 省力化のための変革
 - ▶ 同等の業務をより少ない手数・人員で遂行できれば（何らかの意味で）優位に？

省力化だけが目的ではない

- ▶ 省力化だけが目的ではない
- ▶ だが、それのできる余力により様々な変革が可能だろう
 - ▶ 「顧客」≡「学生・保護者」のニーズに応える
 - ▶ 社会のニーズに応える

国・官公庁の動き

- ▶ 国・官公庁の DX 旗振り
 - ▶ 経済産業省のデジタル・トランスフォーメーション
 - ▶ 企業の DX を促す趣旨のものが多い
 - ▶ 官公庁内部の DX の取り組みもある

官公庁内部の DX のとりくみ

▶ 経済産業省-経産省の新たな挑戦 経産省のデジタル・トランスフォーメーション

「政府全体では行政手続きは年間数億件行われており、その「電子化」はこれまでも進められてきた。しかし、多くの場合は電子化で紙から PC 画面になっただけで、情報の整理、入力、確認など作業の本質は何も変わらず、」

「DX によってどんな変化が今後もたらされるだろうか。異なる手続きで同じ情報を何度も入力する必要があったものが一度の入力で良くなる「ワンスオンリー」や、関連する手続きが一括で終わる「ワンストップ」に。手続きにかかっていた手間・時間が圧倒的に少なくなる。」

文科省による大学等 DX の支援

- ▶ (経産省が企業の DX を促すのと同じ構図).
- ▶ デジタルを活用した大学・高専教育高度化プラン
 - ▶ 「デジタルを活用した大学・高専教育高度化プラン」実施機関の決定について

文科省による大学等 DX の支援（続き）

- ▶ 二つの取り組みについての公募だった
 - ▶ 【取組①】「学修者本位の教育の実現」
 - ▶ （取組例）遠隔授業による成績管理を発展し、学修管理システム（LMS）を導入して全カリキュラムにおいて学生の習熟度等を把握。蓄積された学生の学修ログを AI で解析し、学生個人に最適化された教育（習熟度別学修や履修指導等）を実現
 - ▶ 【取組②】「学びの質の向上」
 - ▶ （取組例）VR（Virtual Reality）を用いた（対面ではない）実験・実習を導入するなど、デジタルを活用して、これまで困難と思われていた内容の遠隔授業を実現。更に、自大学等のみならず、開発した教育システムやデジタルコンテンツ等を他大学等と共有・活用
- ▶ ①は 1 件 1 億円、②は 1 件 3 億円程度の公募

文科省による大学等 DX の支援（続々）

- ▶ 近隣大学だと、金沢工業大学が取り組み①，②それぞれで採択されている
 - ▶ ①「DX による学生一人ひとりの学びに応じた教育実践」
 - ▶ ②「DX による時間と場所の制約を超えた学びの場の創出」

いずれも「教育 DX」の高度化

- ▶ いずれも大学「教育 DX」の高度化支援が目的.
- ▶ 大学「運営 DX」の高度化は？

本学での取り組み

富山大学データサイエンス推進事務室により、DX学修セミナー、同ワークショップ2021が開催されている。

東北大学の取り組み

東北大学 DX ナビゲーション

- ▶ 「オンライン事務化宣言では、「窓口フリー」「印鑑フリー」「働き場所フリー」という3つのフリーを掲げ、その実現のために「業務のDX推進プロジェクト・チーム」を結成しました。」
 - ▶ 「窓口フリー」
 - ▶ 「いつでも誰でもどこからでも、所望する情報に即座にアクセス可能なオンラインカウンターが実現。」
 - ▶ 「印鑑フリー」
 - ▶ 「手続きのオンライン化や電子決裁システム化などの検討・提案を実施。」
 - ▶ 「働き場所フリー」
 - ▶ 「2021年3月までにテレワークに関する規程やマニュアルを整備し、2021年4月、テレワークおよびフレックスタイム制を制度化。」

官公庁の DX を参考にした大学での DX

経産省が自身の DX に取り組んだように、大学も自身の DX にも取り組むべきではないか？ 例えば、

- ▶ 事務手続きの「ワンスオンリー」化
 - ▶ 名前や所属を何度も入力しない、とか
- ▶ 事務手続きの「ワンストップ」化
 - ▶ 出張手続きをしたら
 - ▶ 休講（が必要なら）受講者へのアナウンス
 - ▶ 補講（が必要なら）のアナウンス
 - ▶ 振り替えの休日（が必要なら）の登録
 - ▶ などなど
- ▶ 「大学自身の」どんな DX が考えられるだろうか？

どんな DX が考えられるだろうか？

- ▶ 少し議論の時間をとりたいと思います
- ▶ 本学での DX としてどんなことが考えられるでしょうか？
 - ▶ 「教育 DX」
 - ▶ 「管理運営の DX」
 - ▶ 上述の DX 学修セミナー，ワークショップのお話も伺えたらありがたいです

ここまでの参考文献

- ▶ 「「デジタル・トランスフォーメーション」DXとは何か？ IT化とはどこが違うのか？」
- ▶ 東北大学 DX ナビゲーション
- ▶ 文部科学省デジタルを活用した大学・高専教育高度化プラン

まとめ

- ▶ データサイエンスとは
 - ▶ データを要約する
 - ▶ データの相関を与える
 - ▶ データからデータにないことがらを予測・推定する
 - ▶ 以上のような事柄を目標とする，数学（確率・統計）とデータ処理技術（可視化・計算）
 - ▶ さまざまなサービスがあり，始めるのは簡単！
- ▶ DX とは
 - ▶ デジタル化により社会や生活の形・スタイルを変革すること