

1-5 データ・AI利活用の現場

東京大学 数理・情報教育研究センター

2020年4月28日

2024年6月17日 改訂

概要

- データサイエンティストの仕事の流れとは具体的にどのようなものでしょうか？
- ここではデータサイエンスのサイクルの例として映画製作、流通、製造、金融、サービス、インフラ、公共、ヘルスケア等におけるデータ・AI利活用事例を簡単に見ていきます
- データ・AIを活用することによって、どのような価値が生まれているかを知ることが目標です

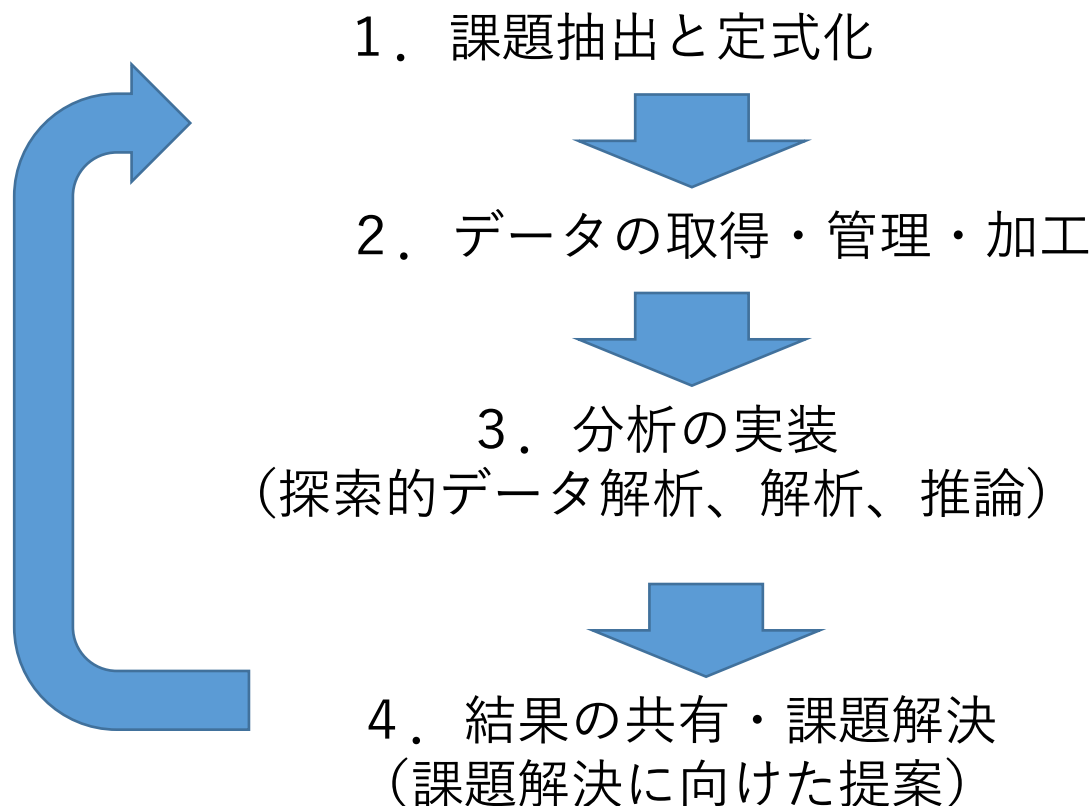
本教材の目次

1. データサイエンスのサイクル	4
2. データ・AI活用事例紹介	10
3. 参考文献	19

1-5-1 データサイエンスのサイクル

データサイエンスのサイクル

- データサイエンスのサイクルは大きく分けると4つのステージに分かれます



Netflixの場合：課題抽出と定式化

- おすすめ（推薦システム）：様々な手法を用います
 - 回帰分析
 - 木モデル（ランダムフォレスト、勾配ブースティング）
 - 強化学習、ニューラルネット
 - 行列分解
 - ファクタリゼーションマシーン
 - 制限付きボルツマンマシン
 - マルコフ連鎖、グラフィカルモデル
 - クラスタリング（K-meansから階層的ディリクレ過程まで）
- 得られた洞察はコンテンツ製作にも役立てられます
 - Netflixは今やどのスタジオよりも映画を製作しています [Bloomberg2019]



Netflixの推薦システム
<https://www.youtube.com/watch?v=0CGQvdAbNcc>
の2:47

データの取得・管理・加工

- ユーザーのウェブログデータを元に趣味・趣向を推定
 - 何を視聴したか、何を繰り返し視聴したか、途中で視聴するのをやめたかなど
- 作品ごとにタグデータの作成
 - 専門のタガーも存在しているとのこと

図は

<https://www.youtube.com/watch?v=0CGQvdAbNcc>

の12:35を参照してください

分析の実装

- 探索的データ解析
 - 得られたデータに対して基本的な分析を行います
- データ解析と推論
 - データ解析の結果を基にして、未知の情報を予測・説明するプロセスを推論と呼びます
- 実験
 - また、分析の結果見込みがあるとされた予想モデルに関してはオンライン実験をすることでその実用性を検証することもあります

多様な趣味・趣向があるようです
<https://www.youtube.com/watch?v=0CGQvdAbNcc>
の9:13

オンライン実験の様子
<https://www.youtube.com/watch?v=0CGQvdAbNcc>
の4:05

結果の共有と課題解決に向けた提案

- 製作段階ごとに人気度を予想したり各要因の効果をインタラクティブな図としてまとめたりすることで製作の意思決定に役立てます
 - 判断支援

製作段階ごとの人気度の予想

<https://www.youtube.com/watch?v=0CGQvdAbNcc>

の13:27

各要因の効果の可視化

<https://www.youtube.com/watch?v=0CGQvdAbNcc>

の16:46

1-5-2 データ・AI利活用事例紹介

教育

- ICTを活用したデジタル学習環境から集められる教育データの活用を通じて教育学習の評価・改善に繋ぐ試みもあります
- デジタル学習環境から集められるログイン履歴や学習コースへのアクセスなどの履歴情報はクリックストリームデータと呼ばれ変化点検知など様々な手法を用いて分析されています[山西 et al.2022]
[Shimada et al.2018]
 - クリックストリームデータは、Webサイトやオンラインショッピングを訪れたお客さんの閲覧履歴などシステムの閲覧履歴一般のことを指します
- そうしたデータ分析だけではなくChatGPTなど生成AIは外国語の学習に活用することもできます
 - <https://dss.i.u-tokyo.ac.jp/blog/chatgpt>で日本語学習を強化する/

芸術

- ニューヨーク近代美術館はRefik Anadol氏の“Unsupervised”（教師なし）という作品を展示しています。この作品は同美術館の200年間の展示物の内、オープンデータになっているものを用い深層学習モデルを訓練し、リアルタイムで新しい美術作品を新規生成し続けるという作品です。これは同美術館の入口に大規模に展示されており、天気や光の具合によって生成される作品が変わるようになっているそうです。
 - 通常の画像の新規生成と違い何か目的をもった美術作品を生成していないことがポイントです。また全ての作品はブロックチェーンで記録しているそうです。
 - <https://www.moma.org/calendar/exhibitions/5535>
- その他の例としては、贋作を検知することにも深層学習は使われています[Mehrjardi et al.2023]

製造と流通

- データサイエンスのサイクルはありとあらゆる分野で使われています
 - 課題解決の中心には必ずデータサイエンティストがいます
- キューピーの食品工場の不良品検知でも様々な課題がありました
 - 食品原料は個体ごとの違いが多く、品種も多種多様です。そのために人手に頼らざるを得なかったのですが、スタッフの作業負荷は非常に高く、労働環境の改善が長らく課題となっていました。
(<https://ai.brainpad.co.jp/case-study/194/>)
- 長距離トラックの自動運転化は、運転手の負担の軽減や運送・物流コストの削減などメリットが大きく注目されていますが、実用化にはまだ様々な課題があると言われています
 - 乗用車より高度な運転技術が必要、開発コストの問題など

金融

- 金融では投資判断においてデータサイエンスの技術が活用されることが多いです
 - 東京大学大学院 情報理工学系研究科では演習講義の一環として、みずほ銀行と一緒に先進的テクノロジーを活用したチャート分析システムの共同開発を行いました
(https://www.i.u-tokyo.ac.jp/news/files/20190926_DSS_final_kouhou.pdf)
- データサイエンスやAI技術で金融サービスの不便を解消しようとするフィンテックも盛んです
 - 金融サービスの拡大
 - 「1-1 社会で起きている変化」で見た通り中国はフィンテックのフロントランナーです

サービス

- データサイエンスによって様々なサービスが改善できます
 - オンラインゲームを開発するコロプラではデータサイエンスは広く活用されています
 - 「スマホゲームの新規ユーザーの定着率が低い」という状況があればその原因を特定し、「ステージが難しい」ことが原因であれば「特に離脱の生じやすいステージの難易度を調整する」などデータサイエンスが広く活用されています
- (https://be-ars.colopl.co.jp/team/data_science2018.html)
- 「1-1 社会で起きている変化」で紹介したチャットボットの導入も問い合わせ業務というサービスの改善を目指したものです

公共政策

- 内閣府や経済産業省がビッグデータを用いて消費動向や販売・生産動向のナウキャストニングを行っていることは「1-3 データ・AIの活用領域」で見た通りです
- 米国の都市開発系の主要シンクタンクであるUrban Instituteでは、データサイエンスをもっと身近にしようと、様々な分析結果をインタラクティブな図としてホームページで公開しています
(<https://www.urban.org/data-viz>)
- データ駆動型の公共政策は世界中で大きな関心を集めています

インフラ

- インフラ業界でもデータサイエンスとAIの利用は広がっています
- トンネルを掘るにはシールド機という特殊な建設機械を用います。こうした特殊な機械は熟練のオペレーターが個人の経験や技量に基づいて運転操作するため、簡単にノウハウが身につくものではないとされていますが、清水建設ではこの建設機械の操作をAIでやらせようと開発を進めています
(<https://www.shimz.co.jp/company/about/news-release/2018/2018005.html>)
- 東京電力が電力需要予測コンテストを開催するなど発電の世界でもデータサイエンスは活用されています
(https://www.tepco.co.jp/press/news/2017/1440911_8963.html)

ヘルスケア

- [Sugiura et al.2018]ではコストのかかる緑内障視野テストの代わりに網膜の厚さのデータを機械学習で分析することで緑内障診断を試みています
(<https://www.youtube.com/watch?v=WtUxFLEr-BQ>)
- Google HealthもAIによる乳がん診断に関する論文を発表しています[Mckinney et al.2020]
- ヘルスケアとAIを融合する研究は今後も益々多くなりそうです

1-5-3 参考文献

参考文献

- [Bloomberg2019] Lucas Shaw, "Netflix Made More Films Than Any Other Studio. Now It Wants an Oscar", Bloomberg, 2019年12月31日 22:00 JST.
- [Mckinney et al.2020] McKinney SM, et al., "International evaluation of an AI system for breast cancer screening", Nature. 2020 Jan;577(7788):89-94. doi: 10.1038/s41586-019-1799-6. Epub 2020 Jan 1.
- [Mehrjardi et al.2023] Fatemeh Zare Mehrjardi, Ali Mohammad Latif, Mohsen Sardari Zarchi, and Razieh Sheikhpour. 2023. A survey on deep learning-based image forgery detection. Pattern Recogn. 144, C (Dec 2023).
- [Shimada et al.2018] Atsushi Shimada, Yuta Taniguchi, Fumiya Okubo, Shin'ichi Konomi, and Hiroaki Ogata. 2018. Online change detection for monitoring individual student behavior via clickstream data on E-book system. In Proceedings of the 8th International Conference on Learning Analytics and Knowledge (LAK '18). Association for Computing Machinery, New York, NY, USA, 446–450.

参考文献

[Sugiura et al.2018] Hiroki Sugiura, Taichi Kiwaki, Siamak Yousefi, Hiroshi Murata, Ryo Asaoka, Kenji Yamanishi, "Estimating Glaucomatous Visual Sensitivity from Retinal Thickness with Pattern-Based Regularization and Visualization", KDD '18: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining July 2018 Pages 783–792, <https://doi.org/10.1145/3219819.3219866>

[山西 et al.2022] 山西健司, 久野遼平, 島田敬士, 峰松翼, 井手剛, 「異常検知からリスク管理へ」, AI/データサイエンス ライブラリ, サイエンス社, 2022.