

ビッグデータとデータエンジニアリング

《学修項目》

- ◎ICT（情報通信技術）の進展、ビッグデータ
- ◎ビッグデータの収集と蓄積、クラウドサービス
- ◎ビッグデータ活用事例（人の行動ログデータ、機械の稼働ログデータ、他）
- 社会で活用されているデータ（種類と所有者）
- ソーシャルメディアデータ、検索エンジンとSNS、オープンデータ

《キーワード》

ビッグデータ、IoT、ロボット、データ量の増加、計算機の処理性能の向上、AIの非連続的進化、第4次産業革命、Society5.0、データ駆動型社会、複数技術を組み合わせたAIサービス、人間の知的活動をAIの関係性、データを起点としたものの見方、人間の知的活動を起点としたものの見方

《参考文献、参考書籍》

- [1] 東京大学MIセンター公開教材 「2-1 ビッグデータとデータエンジニアリング」 《利用条件CC BY-NC-SA》
- [2] 教養としてのデータサイエンス（講談社 データサイエンス入門シリーズ）
- [3] データサイエンスの考え方 社会に役立つAI×データ活用のために（オーム社）
- [4] Pythonによるあららしいデータ分析の教科書（翔泳社）
- [5] オープンデータ活用事例28選とおすすめデータセットまとめ（宙畠） 《利用条件》

1. ICT（情報通信技術）の進展、ビッグデータ

現在、スマートスピーカーや自動運転など、驚異的に発達した人工知能（Artificial Intelligence, AI）を活用した技術が私達の日常生活まで深く入りこんできている。AIの導入は産業界においても積極的に進められており、小売業をはじめとして製造業にいたるまで、ありとあらゆる産業分野でビジネスの形態が変化しつつある。AIが適切に動作するためには、膨大な量のデータ、いわゆる「ビッグデータ」の存在が欠かせない[2]。

ここでは、AIにより引き起こされる社会や日常の大きな変化を認識する。またAIが活用されている領域や現場、代表的な技術などを紹介する。あわせて、AIに必須のデータがどのような形で収集され、いかに整理され分析されるのか、データ処理の大まかな流れについて学ぶ。

1.1 ビッグデータとは / データ量の増大 / ビッグデータ利活用のための課題 [1]

ビッグデータとは

大量の(デジタル)データのこと。多くの情報がデジタル化され、それがインターネットを介して流通することが可能となって来ています。ソーシャルメディアであるソーシャル・ネットワーキング・サービス(SNS)でやりとりされる大量のデータ自体、その利用履歴、インターネット上での購買履歴などは典型的ビッグデータですが、実社会での活動、例えば、コンビニでの購買行動はPOS (point of sale)などの形でビッグデータになっています。気象情報や地震波の情報などの自然現象、加速器による実験データもビッグデータになり得ます。

こうしたデータの中から必要な情報を探し出し、さらには、有効に活用し付加価値をつけることができるようになると期待されています。このような社会的な動きに関連する大量のデータという点を強調して、ここ10年ほど、ビッグデータと呼ばれることが多くなってきました[1]。

[1] <https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/big-data-the-next-frontier-for-innovation>
(Access: 2021/2/24)

東京大学 数理・情報教育研究センター 斎藤洋 2021 CC BY-NC-SA

4

- [Big data: The next frontier for innovation, competition, and productivity | McKinsey](#)

データ量の増大

我々は多くのデータに囲まれています。[1]によれば、毎分、450万YouTubeビデオが視聴され、1億8800万のメール送信がなされ、39万のアプリがダウンロードがされ、Instagramでは約28万のストーリー、5.5万の写真がポストされています(2019年)。

こうした行為がビッグデータの生成を加速しています。

[1]<https://web-assets.domo.com/blog/wp-content/uploads/2019/07/data-never-sleeps-7-896kb.jpg>
Accessed 2021/2/26

東京大学 数理・情報教育研究センター 斎藤洋 2021 CC BY-NC-SA

5

- [DATA NEVER SLEEPS 7.0: How much data is generated every minute? | MONO](#)

ビッグデータ利活用のための課題

ビッグデータ利活用のための課題は以下の4点と言われています。

- ① 膨大なデータの蓄積
- ② 膨大なデータの効率的処理
- ③ 膨大なデータから高付加価値な生成物の作成
- ④ セキュリティ・プライバシー保護

以降のページで、①、②を克服しビッグデータの利活用を実現し得る状況になりつつあることを勉強します。また、その後のページでは③に関わる具体的な事例を述べます。④については、2-6で勉強してください。

*注記：スライド6説明中「2-6」の部分は、本コンテンツではsec.6と読み替えてください。

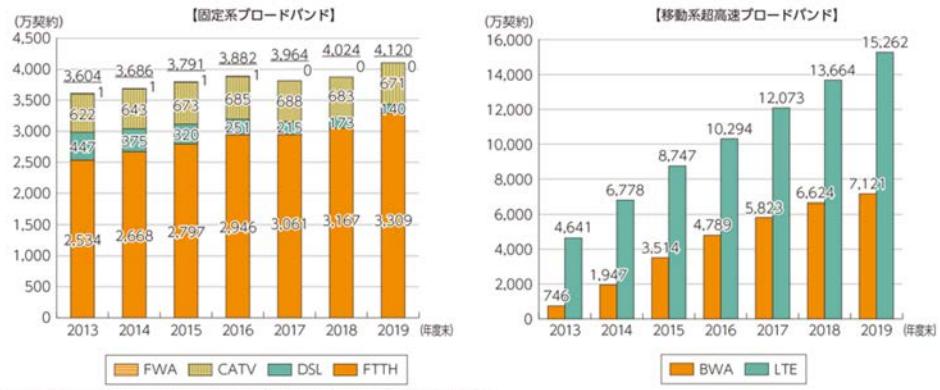
1.2 ビッグデータとデータエンジニアリングを実現可能とした背景：ICT（情報通信技術）の進展 [1]

ビッグデータとデータエンジニアリングを実現可能とした背景：ICT（情報通信技術）の進展

ビッグデータの活用の機運が盛り上がった理由の1つがICT技術の進展であることは間違ひありません。特に、通信技術の発展によりビッグデータ収集能力が飛躍的に向上しました。また、SNSなどの形でデータそれ自体がネットワーク上に出現するようになりました。データ収集に関する具体的な話は2-3で勉強するとして、ここでは、高速な通信ネットワークと携帯電話ネットワークを中心とする無線ネットワークがいかに普及したか、見てみましょう。

ビッグデータとデータエンジニアリングを実現可能とした背景：ICT（情報通信技術）の進展

固定、移動ともに高速化が急速に進展



*過去の数値については、事業者報告の修正があったため、昨年の公表値とは異なる。

FWA: fixed wireless access
CATV: cable television
DSL: digital subscriber line
FTTH: fiber to the home

BWA: broadband wireless access
LTE: long term evolution

出典：総務省ホームページ

図表5-2-2-2 プロードバンド契約数の推移

<https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/r02/html/nd252210.html>

Access: 2021/2/27

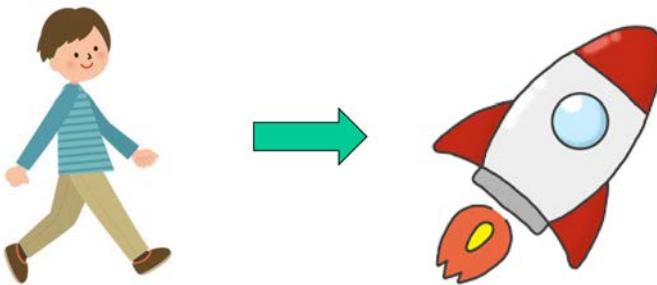
東京大学 数理・情報教育研究センター 斎藤洋 2021 CC BY-NC-SA

8

- 総務省 | 令和2年版情報通信白書 プロードバンド契約数の推移

ビッグデータとデータエンジニアリングを実現可能とした背景：ICT（情報通信技術）の進展

あるインターネット接続サービス事業者の例[1]:
約20年で7800倍の高速化。
これは徒歩とロケットほどの速度差ということです。



[1] <https://www.ocn.ne.jp/ocn20th/history/column1/>
Access: 2021/2/27

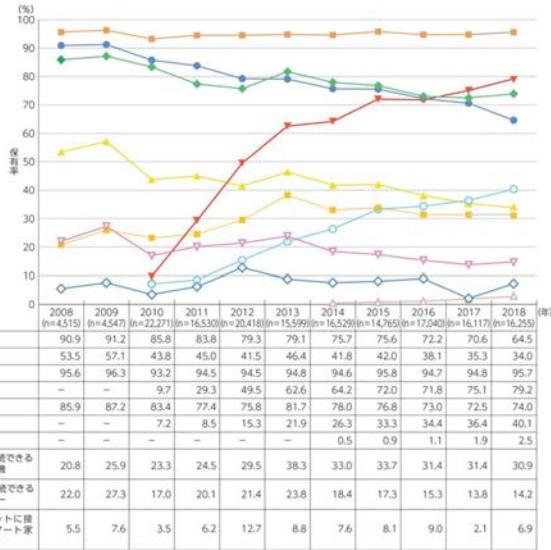
東京大学 数理・情報教育研究センター 斎藤洋 2021 CC BY-NC-SA

9

- OCN | 日本のインターネット人口は20年で1億人を突破 (2016年)

ビッグデータとデータエンジニアリングを実現可能とした背景：ICT（情報通信技術）の進展

- スマホの保有率の急増
- 画像・動画・音声などのupload、downloadの容易化



出典：総務省ホームページ

図表3-2-1-1 情報通信機器の世帯保有率の推移

<https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/r01/image/n3201010.png>

Access: 2021/2/27

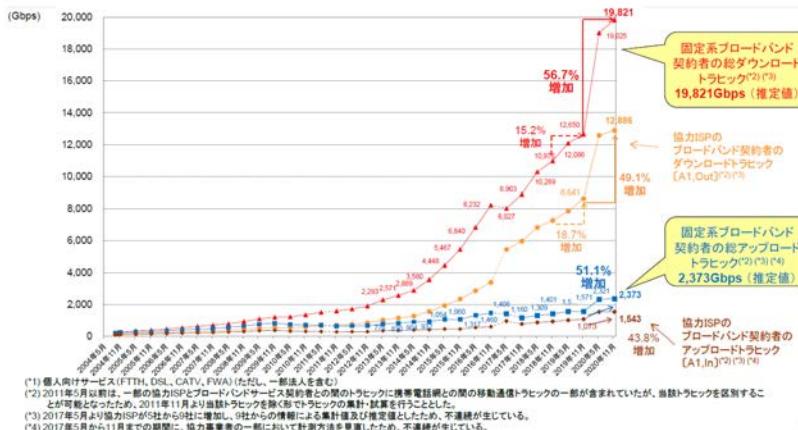
東京大学 数理・情報教育研究センター 斎藤洋 2021 CC BY-NC-SA

10

- 総務省 | 令和2年版情報通信統計 情報通信機器の世帯保有率の推移

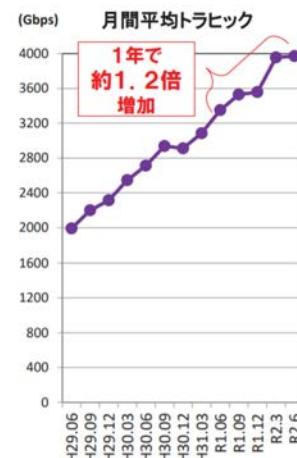
ビッグデータとデータエンジニアリングを実現可能とした背景：ICT（情報通信技術）の進展

急増が続く固定系ブロードバンド契約者のトラヒック（通信量）



ビッグデータとデータエンジニアリングを実現可能とした背景：ICT（情報通信技術）の進展

移動体通信トラヒックも大幅に増加



出典：総務省ホームページ
移動体通信トラヒックの推移
<https://www.soumu.go.jp/johotsusintokei/field/da/gt010602.pdf>
Access: 2021/2/27

東京大学 数理・情報教育研究センター 斎藤洋 2021 CC BY-NC-SA

12

- 総務省 | 令和3年版情報通信統計 移動体トラヒックの推移

*注記：スライド7説明中「2-3」の部分は、本コンテンツでは sec.3 と読み替えてください。

2. ビッグデータの収集と蓄積、クラウドサービス

ビッグデータの特徴として、莫大な量を持つことであることは明らかであるが、それ以外にもいくつかの特徴がある[2]

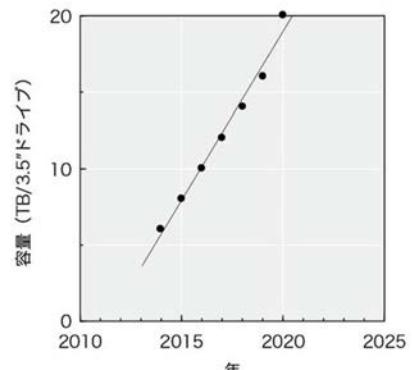
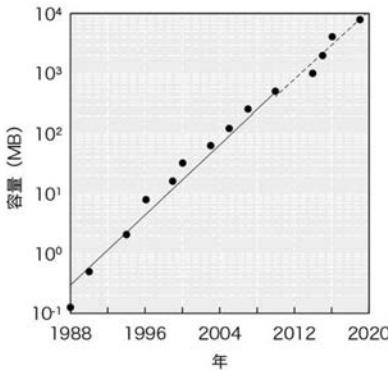
ビッグデータの代表的な特徴は、英語の頭文字をとって 3V といわれる：(1) Volume (量) - データの量が大きいこと, (2) Velocity(速度) - データが計測され、記録されるスピードが速いこと, (3) Variety(種類) - データの種類がさまざまであること。

データが計測・記録されるスピードも桁違いに増している。スマートスピーカー(smart speaker)と呼ばれる製品群(Google Home や Amazon Echo)は、世界中の各家庭の生活者の質問を秒単位で集積し解析している。(3V の Velocity)。データの種類が多様である(3V の Variety)ことも、ビッグデータの特徴の1つである。ビッグデータの種類については後述する。

2.1 ビッグデータとデータエンジニアリングを実現可能とした背景：ストレージ、演算速度、GPGPU、クラウドコンピューティング [1]

ビッグデータとデータエンジニアリングを実現可能とした背景：ICT（情報通信技術）の進展

ビッグデータ利活用のための課題の1つである膨大なデータの蓄積をICT技術が解決しつつある証左として蓄積容量の拡大のデータを示します。また、これらを用いたデータベース技術については、2-4で勉強します。



出典 NHK [1] <https://www.nhk.or.jp/strl/publica/rd/181/2.html>
Access: 2021/3/26

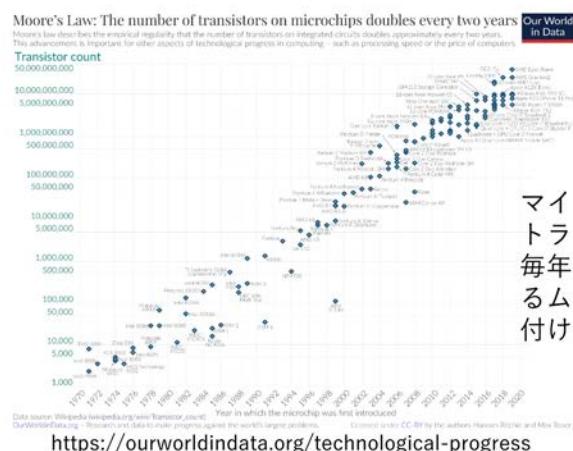
東京大学 数理・情報教育研究センター 斎藤洋 2021 CC BY-NC-SA

13

- NHK放送技術研究所 | ストレージデバイスの研究開発動向 (2020年)

ビッグデータとデータエンジニアリングを実現可能とした背景：ICT（情報通信技術）の進展

ビッグデータ利活用のための課題のもう1つは、膨大なデータの効率的処理です。プロセッサ速度の向上に関するデータから、この課題を克服しつつあるということがうかがえます。



マイクロチップ上のトランジスター数は毎年2倍になるとすむムーアの法則を裏付けるグラフ

Data source: Wikipedia (https://en.wikipedia.org/wiki/Transistor_count).
OurWorldInData - Research and data to make progress against the world's largest problems.
<https://ourworldindata.org/technological-progress>

Access: 2021/2/26

東京大学 数理・情報教育研究センター 斎藤洋 2021 CC BY-NC-SA

14

- Our World in Data | Moore's Law (2020年5月)

ビッグデータとデータエンジニアリングを実現可能とした背景：ICT（情報通信技術）の進展

さらに、以下のような技術シーズがビッグデータの活用を後押ししています。

1つは、GPU(graphics processing unit)です。GPUとは、画像処理用のプロセッサのことです、ゲーム機などに搭載されてきました。その後、それがCADや科学計算にも使われるようになりました。これをGPGPU(General-purpose computing on graphics processing units)と言います。データサイエンスやAIで使われるニューラルネットワークなどの機械学習アルゴリズムでもGPUの利用を想定して高速化を図っているものもあり、それらはGPUがないコンピュータでは動作しません。最近では、多くのコンピュータにGPUが搭載されています。

ビッグデータとデータエンジニアリングを実現可能とした背景：ICT（情報通信技術）の進展

さらに、クラウド（コンピュータクラウド）と呼ばれる技術もビッグデータの活用を後押ししています。

クラウド（コンピュータクラウド）とは、コンピュータ群をネットワークの先のネットワークセンターに用意し、ユーザは、そのコンピュータ群（の一部）を必要に応じて利用するような環境のことです。GPUを具備するようなコンピュータなどを専用に用意して、データサイエンスやAIを行うことは、初期投資などの点から、ハードルが高いですが、クラウドの形態であれば、容易に実現できます。これも、ビッグデータを用いて機械学習が身近になって行われる理由の1つです。国内のIT企業やGoogleやAmazon, Microsoftなどの海外の有力企業がクラウドサービスを提供しています。

*注記：スライド13説明中「2-4」の部分は、本コンテンツではsec.4と読み替えてください。

2.2 事業活動におけるデータ・AI活用の広がり [2]

データとAIは、さまざまな事業活動の中で活用されている。企業における事業活動は、バリューチェーン(value chain)（図2.2.1）にあてはめて整理することができる。ここではバリューチェーンの研究開発、購買物流・調達、製造、出荷物流、マーケティング、販売、サービスにおいて、どのようにデータとAIが活用されているのか、複数の活用例を紹介しながら解説する。

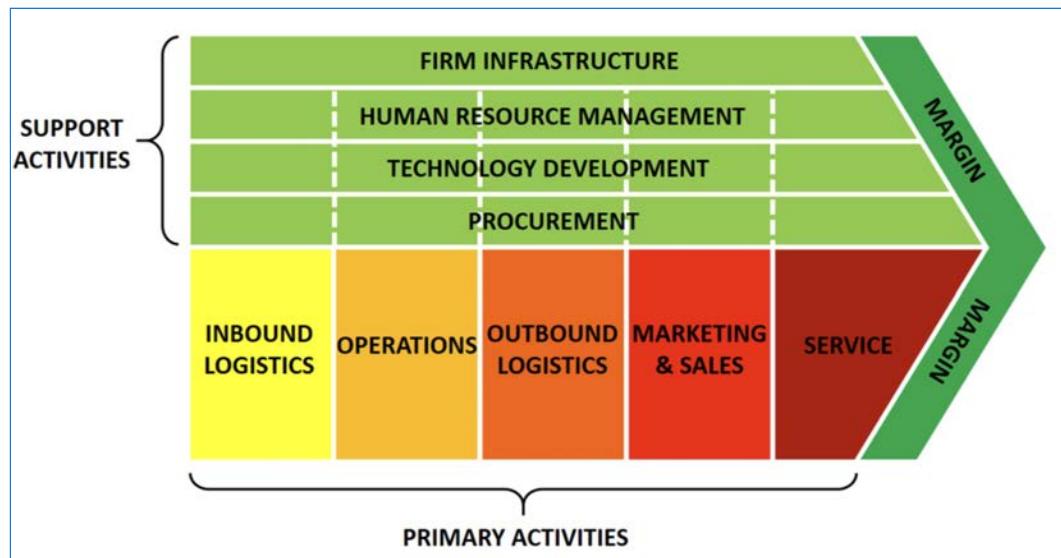


図2.2.1 出所：B2U 公式Webサイト | Value Chain Analysis: An Internal Assessment of Competitive Advantage

2.2.1 研究開発におけるデータ・AI利活用 [2]

研究開発の領域では、技術開発や素材開発などにデータとAIが活用されている。

たとえば自動車開発においては、自動運転に向けた技術開発が行われている。自動車の位置情報や地図情報画像認識技術やセンシング技術（図2.2.2）を活用し、人の運転操作なしに走行できる自動車の開発を目指している。



図2.2.2 出所：自動運転ラボ 公式Webサイト | 自動運転で扱うデータ群まとめ センサーデータや位置情報、地図 (2021年)

素材開発においては、過去の実験データやシミュレーションデータから新素材を探索する取り組みが始まっている。この新素材開発に関する取り組みはマテリアルズインフォマティクス（図2.2.3）と呼ばれ、素材開発期間の短縮や未知の素材発見につながる可能性があると期待されている。



【従来技術との比較表】

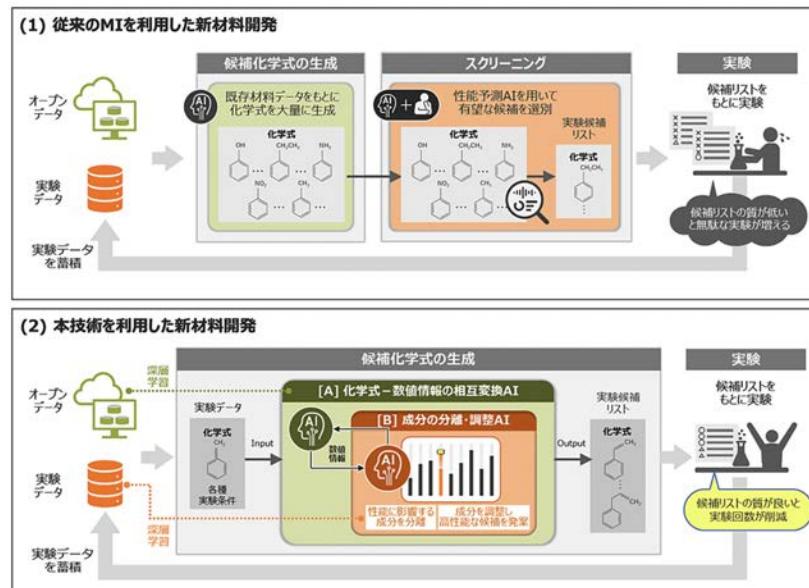


図2.2.3 出所：三井化学 公式Webサイト | 材料開発を高速化するAI技術の実用化に向けた実証試験を開始(2021年)

2.2.2 購買物流・調達におけるデータ・AI利活用 [2]

購買物流・調達の領域では、**在庫適正化や調達コスト削減（図2.2.4）**（不要な在庫を抱えすぎないために、製造する製品の需要予測を行い、部品や原材料の発注量を最適化する取り組み）などにデータとAIが活用されている。

物流の基礎知識と現場改善が学べるサイト
物流効率化のキホン

HOME > 改善事例・学ぶ・調べる > 物流効率化のキホン

基礎から学ぶ
物流現場の効率化

PDFで詳しく見る

物流の基礎 >

- ▶ 物流の基礎と目的について
- ▶ 物流の歴史
- ▶ これからの物流

物流の役割 >

- ▶ 物流の種類(領域)
- ▶ 物流の機能(輸送・配送・運送)
- ▶ 物流の機能(保管・荷役)
- ▶ 物流の機能(流通加工・包装)
- ▶ 物流の機能(情報システム)
- ▶ 物流設備(マテハン機器)の役割
- ▶ 物流センターの種類

図2.2.4 出所：キーエンス 公式Webサイト | 物流効率化のキホン, 物流現場の効率化

原材料の生産量や調達価格が変動する場合、市場の動向を踏まえ適切なタイミングで取引する必要がある。原材料の価格を予測し、適切なタイミングで取引を行うことで、調達コストを削減しなければ

ならないが、このときの調達先の選定や調達価格の決定（自動最適化）（図2.2.5）においても、データとAIが活用されている。



図2.2.5 出所：GRID 公式Webサイト | ReNom SCM - AI生産計画アプリ

2.2.3 製造におけるデータ・AI利活用 [2]

製造の領域では、品質管理や歩留まり改善などにデータとAIが活用されている。検査工程の省力化や生産計画の最適化にも、データとAIが活用されている。

たとえば画像認識技術を用いて、これまで目視で行っていた検査工程を自動化（図2.2.6）する取り組みが進められている。カメラで撮影した製品画像から、良品と不良品を判別し、自動的に不良品を検出する。



図2.2.6 出所：パナソニック 公式Webサイト | AI外観検査ソリューション「WisSight」

また、これまで熟練者が実施していた生産計画を自動化（図2.2.7）する取り組みも行われている。



図2.2.7 出所：日立 公式Webサイト | Hitachi AI Technology / 計画最適化サービス

2.2.4 出荷物流におけるデータ・AI利活用 [2]

出荷物流の領域では、輸送業務の効率化や倉庫内作業の効率化などにデータとAIが活用されている。

物流では、トラックや鉄道などによる貨物の輸送と、物流拠点での貨物の荷役・保管が行われる。輸送業務では、需要予測を行うことで必要な輸送量を把握（図2.2.8）し、適切な数のトラックやドライバーを手配する取り組みが行われている。

mylogi

03-6304-2541
営業時間平日：10:00~19:00

資料請求・お問い合わせ

コラム | 機能一覧 | mylogiの特徴 | 会社概要 | マイロジ利用料について | 各種ダウンロード | お問い合わせ

mylogi > 物流 > AIが解決する物流の課題！ヒューマンエラーの予防やビジネスの最適化に役立つAIについて解説します。

AIが解決する物流の課題！ヒューマンエラーの予防やビジネスの最適化に役立つAIについて解説します。

© 2020年11月26日 | 物流 | 物流



AIが解決する物流の課題！
ヒューマンエラーの予防やビジネスの最適化
に役立つAIについて解説します。



Last Updated on 2021年12月21日 by art-mylogi

物流業は様々な業務があり、また倉庫以外の届け先なども現場となる複雑な業界です。また、ECサイトの利便性の高まりに伴って、いくつかの課題も生じています。そんな物流業において、今注目されているAIはソリューションとなりうるのか、また物流AIの今後の課題は何なのかを、具体例やデジタルトランスフォーメーションで解説します。

ECサイト事業主様へ！
動画で解説
物流システム
mylogi

図2.2.8 出所：mylogi 公式Webサイト | AIが解決する物流の課題 ヒューマンエラーの予防やビジネスの最適化に役立つAI

物流拠点では、[倉庫内の棚の配置を最適化（図2.2.9）](#)することによって、作業員のピッキング業務を効率化する取り組みが行われている。また、画像認識技術を活用して、倉庫内の商品を自動的に仕分けする取り組みも進められている。

トップ | デジタルアニーラとは | サービス | 量子インスピアード技術の活用 | 技術解説・論文 | お問い合わせ

株式会社富士通ITプロダクツ様
倉庫内の部品集約作業における移動距離を最大45%短縮
最適な部品集約作業の動線を瞬時に導き出す、デジタルアニーラ



量子現象に着想を得たデジタル回路によって組合せ最適化問題を高速に解くことができる、富士通のデジタルアニーラ。
ICT技術の発達とAIの実用化に伴い、これまで以上に複雑な計算を瞬時に行うコンピュータが求められている昨今、製造、流通、金融など様々なビジネス分野での活用が期待されています。今回はデジタルアニーラを導入し、倉庫内作業の効率を劇的に向上させることに成功した富士通ITプロダクツ様の事例を紹介します。

▶ 動画を見る

効率化を追求する工場とデジタルアニーラの出会い

図2.2.9 出所：富士通 公式Webサイト | デジタルアニーラ 倉庫内の部品集約作業における移動距離を最大45%短縮

2.2.5 マーケティングにおけるデータ・AI利活用 [2]

マーケティングの領域では、商品推薦や価格設定などにデータとAIが活用されている。

たとえばECサイトでは、商品の購入履歴やサイト閲覧履歴から、興味・関心がありそうな商品を推薦するレコメンデーションが行われている。また、収益の最大化を目的に、**需要と供給に応じて価格を変動させるダイナミックプライシング（図2.2.10）**に関する取り組みも進められている。

The screenshot shows the homepage of Alsmiley, a portal for AI products and services. The main navigation bar includes links for 'AIニュース' (AI News) and 'AIイベント' (AI Events). A featured news article titled 'AIの進展でダイナミックプライシングがより活発化！国内外での活用事例' (Dynamic Pricing becomes more active due to AI progress! Examples of use cases abroad) is displayed. The article includes a large image of a city skyline with a superimposed candlestick chart and a yellow arrow pointing upwards, symbolizing growth. Below the image, a brief summary discusses how AI is used to change prices based on demand and supply.

図2.2.10 出所：Alsmiley 公式Webサイト | AIの進展でダイナミックプライシングがより活発化 国内外での活用事例

顧客獲得や顧客育成にも、データとAIが活用されている。新たな顧客を獲得するために、従来からあるマスメディア（テレビ、新聞、雑誌、ラジオなど）に加え、**検索連動型広告（図2.2.11）**などのインターネット広告の活用が進んでいる。検索連動型広告では、検索エンジンでユーザが検索したキーワードに関連する広告が表示される。

The screenshot shows the homepage of Bigmac inc., a company specializing in marketing and SEO. The top navigation bar includes links for 'TOP', '企業情報', '事業内容', '事例', 'ブログ', 'コラム', 'お知らせ', 'お問い合わせ', '資料請求', and '採用情報'. A sidebar on the right features a 'お問い合わせ' (Contact Us) section with a live chat interface and a 'SEARCH' bar. The main content area displays a blog post titled '検索連動型広告とは？ディスプレイ広告との違いも紹介' (What is search-driven advertising? Also introduce the difference from display ads). The post includes a small illustration of a person pointing at a chalkboard. To the right of the post, a 'RANKING' section lists four articles with their titles and publication dates:

- 1. 閲覧履歴が残らないシークレットモード！メリットとデメリットを紹介 (Published 2018/06/01)
- 2. Google予測変換が邪魔！端末・ブラウザごとの削除方法を解説！ (Published 2020/04/07)
- 3. PNGとJPEGの違いは？あなたの画像に最適な保存形式を！ (Published 2019/06/28)
- 4. Googleスプレッドシートで使える日付の範数・シリアル化をご紹介 (Published 2016/06/28)

図2.2.11 出所：Bigmac inc. 公式Webサイト | 検索連動型広告とは？ディスプレイ広告との違い

2.2.6 販売におけるデータ・AI利活用 [2]

販売の領域では、[販売管理や予算管理などにデータとAIが活用（図2.2.12）](#)されている。

たとえば流通業では、POS(Point of Sale)システムを導入し、レジでバーコードをスキャンすることによって販売実績を蓄積している。蓄積された販売実績は、週次や月次で実施する売上分析に利用される。売上分析では、販売実績の増減傾向や予算達成状況などを確認する。

The screenshot shows a news article from Google Cloud's website. The title is "株式会社セブン-イレブン・ジャパン：これからの中戦略を支えるデジタルデータ基盤「セブンセントラル」をGoogle Cloud上に構築". Below the title are the logos for Google Cloud and Seven-Eleven Japan. The main text discusses the company's digital transformation strategy and its implementation on Google Cloud. It mentions the challenges of legacy systems and the need for DX (Digital Transformation). The author is the Google Cloud Japan Team, dated September 9, 2020. A "Try Google Cloud" button is also present.

図2.2.12 出所：Google Cloud 公式Webサイト | 株式会社セブン-イレブン・ジャパン：デジタルデータ基盤「セブンセントラル」

また営業活動の効率化にも、データとAIが活用されている。日々の営業活動が記録された営業日報を分析することによって、効率的な営業活動につなげる取り組みが行われている。営業日報は、テキスト（文書）で記録されているため、[自然言語処理技術を用いて分析（図2.2.13）](#)が行われる。また、営業日報から提案の進捗状況を把握し、最適な訪問計画を作成する取り組みも行われている。



図2.2.13 出所：Stockmark 公式Webサイト | テキストデータマイニングソリューション

2.2.7 サービスにおけるデータ・AI利活用 [2]

サービスの領域では、修理やメンテナンスなどのアフターサービスにデータとAIが活用されている。たとえば自動車や工作機械は、安定的に稼働させるために、定期的な修理・メンテナンスが必要になる。設備の故障が生産ラインの停止や重大な事故につながる可能性がある場合、設備を常時監視し、異常があれば早期に復旧することが求められる。そのため稼働記録やセンサデータをもとに異常を検知（図2.2.14）し、故障している箇所を特定する取り組みが行われている。また、設備の劣化を予測し、故障前に修理交換を行う予防的メンテナンスに関する取り組みも進められている。

The screenshot shows an article from the ASPIック website. The header includes the logo, the text "ASPIック 公式サービス", and navigation links for "無料レポート", "ジャンルから探す", and "特集記事". The breadcrumb navigation shows "TOP > 特集記事 > 異常検知ソリューション8選。タイプ別に紹介". The main title is "異常検知ソリューション8選。タイプ別に紹介". Below the title is a photo of a man working on a laptop. The text "最終更新日:2022-07-29" is at the bottom right. The content area has a blue sidebar with a table of contents: "▽ 目次" followed by a list of topics under "▼ 異常検知ソリューションとは？", "▼ 異常検知ソリューションのタイプ", "▼ 主な異常検知ソリューション（機械・設備監視向け）", "▼ 主な異常検知ソリューション（システム監視向け）", "▼ 主な異常検知ソリューション（カメラ解析向け）", and "▼ 主な異常検知ソリューション（異音検知向け）".

図2.2.14 出所：アスピック 公式Webサイト | 異常検知ソリューション8選 タイプ別に紹介

顧客対応の高度化や効率化にも、データとAIが活用されている。コールセンターでは、[オペレータ業務を効率化するために、音声認識技術や検索技術の活用が進んでいる（図2.2.15）](#)。また、顧客からの問い合わせ対応をチャットボットによって自動化する取り組みも始まっている。

The screenshot shows a web page from the Konanavi website. The header includes the logo 'コネナビ' (powered by WILLOF), navigation links for '記事', 'インタビュー', and 'コネナビとは', and buttons for 'お問い合わせ' and 'お役立ち資料'. The main content is an article titled 'AI音声認識でコールセンター業務効率化！活用事例や選定ポイントを解説' (Published on June 28, 2022). The article features a large image of a robot interacting with mobile devices, with the text 'AI VOICE RECOGNITION AI音声認識でコールセンター業務効率化' and checkboxes for '活用事例' and '選定ポイント'. To the right, there's a sidebar titled 'よく読まれている記事' with three recommended articles and a 'お役立ち資料' button. Below the article, there's a 'カテゴリー' section with various tags like 'システム', '人材採用', '品質', etc.

図2.2.15 出所：コネナビ 公式Webサイト | AI音声認識でコールセンター業務効率化 活用事例や選定ポイントを解説

3. ビッグデータ活用事例

様々な場面でのビッグデータの活用事例について紹介する。具体的には、POSの活用、電子商取引市場(EC)、推奨行為事例(レコメンド)、人流測定、機械の稼働ログ、ICTシステムログ、画像・音声・文書・自然言語処理、ならびに産業分野別(農業、漁業、製造業、エネルギー産業、医療、インフラ管理など)における事例である。

3.1 POSの活用、電子商取引市場(EC)での事例 [1]

ビッグデータの活用事例：POSの活用

お店に行って商品を購入した際の記録は、POS (point of sale)などと呼ばれ、マーケティングにおいて重要なデータとなります。例えば、どの商品とどの商品は同時に購入される可能性が高いかが分かれれば、それらを近くに並べる商品陳列にすることにつながります。どの曜日にどんな商品が売れるかが分かれれば、曜日に応じた仕入れの量を変えることができます。天候やイベント情報と組み合わせて分析できれば、それらに対応した仕入れ量の決定も可能です。[1]にある外食産業の事例では、POSデータからキャンペーン効果を分析することにより、広告費を削減しながら売り上げの増加を達成しました。これらは実市場でのPOSですが、インターネット上での購買行動はより詳細なデータの取得が可能となります。それらを用いることが、より増加すると想定されます。

[1]<https://liskul.com/one-to-onemarketing-13202>

- LISKUL | One to Oneマーケティングとは？成功例と4手法まとめ(2019年)

ビッグデータの活用事例：ECでの事例

インターネット上の行動は、一人一人の行動ログ（記録）という形で記録に残ります。特に、自社サイト利用者のCookieなどの情報や電子商取引市場(EC)での購買行動履歴を利用する立場にある事業者は、それを用いて顧客ごとのマーケティングを実行できるようになりました。これを従来のマスマーケティングとの対照でOne-to-oneマーケティングと呼ぶ人もいます。具体的には以下のようなことが可能になりました。

●推薦行為

いわゆる「おすすめ商品」の提示です。

●関連情報提供行為

関心がありそうな行動をするユーザに、興味を持ちそうな、購買につながりそうな情報を再提示する、あるいは、新たに提示することです。

3.2 推奨行為事例（レコメンド） [1]

ビッグデータの活用事例：推薦行為事例

ビッグデータ活用の成功例としてもよく取り上げられる推薦行為（レコメンド、リコメンドなどと呼ばれることが多い）について具体的に見てみましょう。用いるデータの種類から、2つに大別されます。

1) 多数のユーザの過去の購買履歴による方法

ユーザ嗜好を多数のユーザの過去の購買履歴に基づきモデル化するものです。ソーシャルフィルタリング、協調フィルタリングなどとも呼ばれています。似た購買履歴ユーザの購入商品から推薦を行えるため、年齢や性別などのユーザ属性や種別などの商品属性が不要な反面、新しい商品など購買履歴がないものに対する推薦ができません。属性に直接リンクしない要因も反映できます。

2) 属性ベースの方法

ユーザ属性や商品の属性を利用する方法。属性情報を得られない場合には適用できませんが、履歴データのない新商品などにも使えます。

ビッグデータの活用事例：推薦行為事例

推薦行為の具体例について見てみましょう。まずは、推薦行為の発祥ともいえるAmazon.comの事業例です。



ビッグデータの活用事例：推薦行為事例

推薦行為の具体例について見てみましょう。まずは、推薦行為の発祥ともいえるAmazon.comの事業例です。

Amazon.comでは、推薦システムが、単に商品を推薦するのみならずWebサイトをパーソナライズするために用いられています[1]。Item-to-Item Collaborative Filteringと呼ばれる推薦アルゴリズムが1998年に導入されました[2]。購入商品と類似すると判定された商品を推薦します。類似商品表作成など、計算の大部分をオフラインで実施しておき、推薦対象ユーザの購入物をもとに表検索を行う部分だけオンラインで計算することで高速に推薦をおこなっています[1]。その結果Amazon.comのサイトの視聴の30%は、推薦に基づくものと推定されています[2]。

[1] Greg Linden, Brent Smith, and Jeremy York, "Amazon.com Recommendations Item-to-Item Collaborative Filtering," IEEE Internet Computing, 7, 1, pp. 76-80, 2003.

[2] Brent Smith and Greg Linden, "Two Decades of Recommender Systems at Amazon.com," IEEE Internet Computing, 21, 3, pp. 12-18, 2017.

- Greg Linden, Brent Smith, and Jeremy York, "Amazon.com Recommendations Item-to-Item Collaborative Filtering," IEEE Internet Computing, 7, 1, pp. 76-80, 2003.
- Brent Smith and Greg Linden, "Two Decades of Recommender Systems at Amazon.com," IEEE Internet Computing, 21, 3, pp. 12-18, 2017.

ビッグデータの活用事例：推薦行為事例

推薦行為の別の具体例としてYouTubeについて見てみましょう。

YouTube の推薦アルゴリズムは以下のよう�습니다[1]。あるビデオAを見た人が見る別のビデオ（共通視聴数）に基づく関連性（relatedness）を尺度にビデオAの関連性Top-Nビデオ(Aの関連ビデオ集合R(A)) をあらかじめ計算しておきます。そしてユーザの視聴履歴などとこの計算結果の表を組み合わせて推薦ビデオを提示しています。例えば、Aを見たユーザには「R(A)に含まれるビデオ」と「R(A)に含まれるビデオの関連ビデオ集合のビデオ」、などを提示するという具合です。

[1] J. Davidson, et al., "The YouTube Video Recommendation System," Proc. 4th ACM Conf. Recommender Systems, pp. 293–296, 2010.

- J. Davidson, et al., "The YouTube Video Recommendation System," Proc. 4th ACM Conf. Recommender Systems, pp. 293–296, 2010.

ビッグデータの活用事例：推薦行為事例

Netflixの推薦はどうなっているのでしょうか。

Netflixの80%は推薦によると言われています[1]。それほど推薦の影響は大きいわけです。現在Netflixの推薦は、多くのアルゴリズムの集合体になっており、画面の下部にマトリクス状に配置された小画面の各行が1つのアルゴリズム（例えば、特定のカテゴリの全ビデオについて、そのユーザの属性から予想される評価点順に提示する）に対応しています[2]。

[1] Brent Smith and Greg Linden, "Two Decades of Recommender Systems at Amazon.com," IEEE Internet Computing, 21, 3, pp. 12-18, 2017.

[2] C.A. Gomez-Uribe and N. Hunt, "The Netflix Recommender System: Algorithms, Business Value, and Innovation," ACM Trans. Management Information Systems, 6, 4, pp. 1-19, 2016.

- Brent Smith and Greg Linden, "Two Decades of Recommender Systems at Amazon.com," IEEE Internet Computing, 21, 3, pp. 12-18, 2017.
- C.A. Gomez-Uribe and N. Hunt, "The Netflix Recommender System: Algorithms, Business Value, and Innovation," ACM Trans. Management Information Systems, 6, 4, pp. 1-19, 2016.

3.3 人流測定, 機械の稼働ログ, ICTシステムログ [1]

ビッグデータの活用事例：人流測定

人流測定による行動把握と予測、制御もビッグデータの有力な活用分野です。例えば、

- 博物館内の人流れを測定し、見学ルートを変更する[1]
- 店内の商品レイアウトを変更する
- 道案内標識の適切な設定位置の決定
- 電子表示装置と連動させた避難ルート指示[2]

などがあります。

[1] <https://xtech.nikkei.com/it/atcl/column/14/122600137/122600001/> [Access 2021/2/26]

[2] <https://www.airc.aist.go.jp/achievements/ja/p-028.html>

通常時と災害時で電子表示装置の表示内容を変化させる



[2] (産業技術総合研究所人工知能研究センター社会知能研究チーム)

- 日経XTECH | 格差広げるビッグデータ100 Part1 先進事例 知らぬ間に不満を解消 (2015年)
- 人工知能研究センター | 人流計測+シミュレーション (2014年)

ビッグデータの活用事例：機械の稼働ログ

人の行動ログもビッグデータとなったように機械の稼働ログもビッグデータとなり得ます。その先駆けとも言える事例がコマツの建設機械に標準装備され稼働状況を監視するKOMTRAXです。インターネットなどのネットワークを通じ、機械のデータがサーバに送信されます。これにより、世界中のコマツの機械の状況が一元的に把握できます[1]。

また、IHIのガスタービン発電プラント運用支援ではプラントの運転データを収集し蓄積し、集まったビッグデータをもとに故障の予兆を発見します[2]。

[1]<https://sanki.komatsu/komtrax/>

[2]https://www.ihi.co.jp/powersystems/lifecycle/operation_support/index.html

- コマツ | 産機KOMTRAX 稼働管理システム
- IHI原動機 | ガスタービン発電プラントの運転状況見守り

ビッグデータの活用事例：情報通信システムログ

情報通信システムのログもビッグデータです。これをもとに、故障診断などを行う試みがなされています[1]。

もともと大規模ネットワークには、事前設定されたルールに従いアラームが発生する仕組みが備わっています。しかし、ハードウェアの致命的故障など単体で異常と分かるものの以外のもの、一時的な故障や、致命的な故障に至る段階での軽微な故障には、十分には機能せず、ベテランオペレータの経験によるところも多いのが実際です。そこで、故障に関わるメッセージからコンソール履歴まで含むログデータの活用が期待されます。

故障などの1つのイベントが発生すると関連する多くの機器で大量のメッセージが生成されます。こうした共起関係を利用して何が起きているかを把握することができるようになり故障診断精度が上がります。

[1]木村達明、Syslog分析による大規模ネットワーク故障診断、電子情報通信学会誌、98、9、pp.823-828、2015.

- 木村達明, Syslog分析による大規模ネットワーク故障 診断. 電子情報通信学会誌, 98, 9, pp.823-828, 2015.(PDF 再配布禁止)
- 中電 技術開発ニュース | ログデータ分析による通信ネットワーク機器の異常検出技術 (2020年)

3.4 活用データ別：画像、音声、文書、自然言語 [1]

ビッグデータの活用事例：活用データ別

これ以外にもビッグデータを機械学習・AIなどで処理する事例は数多く見受けられます。各分野について見ていきましょう。

1. 画像データ活用

画像データは非常に活用が進んでいるようです。

- 自動運転車両プログラム[1]
- 医療画像データからのガンなどの検出[2][5]
- 食品原料検査[3]
- 車両画像識別[4]

[1]<https://jp.techcrunch.com/tag/mobileye/>
[2]<https://pubmed.ncbi.nlm.nih.gov/21452728/>
[3] <https://monoist.atmarkit.co.jp/mn/articles/2004/03/news080.html>
[4] <https://cloudplatform-jp.googleblog.com/2017/05/googlecloudplatform-aucnet-ibs.html>
[5] <https://aitimes.media/2019/06/21/3013/>
[1] -[4] Access 2021/3/6, [5] Access 2021/3/7

- TechCrunch | How Google's self-driving car project accidentally spawned its robotic delivery rival (2021年)
- The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI) (Med Phys. 2011)
- MONOist | AIで原料を判定、食品業界全体にAI活用拡大を目指すキユーピーの挑戦 (2020年)
- Google Cloud 顧客事例 | 株式会社オークネット・アイビーエス 「Konpeki」 ビッグデータ解析 (2017年)

ビッグデータの活用事例：活用データ別

2. 音声データ

音声は認識に加えて、最近では、合成も盛んです。

●音声認識

Apple 「Siri」
Google 「Google assistant」
NTTドコモ 「しゃべってコンシェル」
Amazon 「Amazon Echo」
Microsoft 「Cortana」, 「Azure cognitive services」
など

●音声合成

ニュース読み、AIアナウンサー
[1]<https://www.itmedia.co.jp/news/articles/1803/26/news127.html>
[2]<https://www.nhk.or.jp/voice/yomiko/>
[3]<https://www.sony.jp/professional/ai-announcer/>
[4]<https://www.nikkei.com/theme/?dw=19102801>
[1]-[4]Access 2021/3/6

- ITmedia NEWS | NHKの人造アナウンサー「ニュースのヨミ子」さん (2018年)
- SONY | AIアナウンサー「荒木ゆい」 (2020年)

- 日本経済新聞 | 最新ニュースを2分程度の動画にまとめてお届け (随時更新)

ビッグデータの活用事例：活用データ別

3.文書データ、自然言語データ

翻訳に加えて、最近では、文章の合成や原稿作成が行われるようになってきました。

●機械翻訳

Google 「Google natural machine translation」

Microsoft 「Microsoft translation」, 「Azure」

など

●文書合成、原稿作成

Articoolo

PSW

ライターズボックス

EasyWriter

automatedinsights

など

- DEEPL翻訳ツール
- ライカツ | 日本語の文章自動生成サイト・アプリ・ツール10選 (2022年)

3.5 産業分野別：農業、漁業、製造業、エネルギー産業、医療、インフラ管理 [1]

ビッグデータの活用事例：産業分野別

産業分野別で見てみましょう。

1. 農業

●圃場の状況を撮影したり、センサーで計測したりして集めたビッグデータを解析し、効率的に栽培管理し、ブドウの生産に用いる[1]-[3]

●ビッグデータを使って農業の生産性の向上や経営の改善を目指す農業データ連携基盤協議会 -WAGRI協議会- の発足[4]

[1]<https://smartagri-jp.com/smartagri/20>

[2]<https://smartagri-jp.com/smartagri/113>

[3]<https://smartagri-jp.com/smartagri/155>

[4]<https://wagri.net/ja-jp/>

[1]-[4]Access 2021/3/6

- SMART AGRI
- 農業データ連携基盤（通称：WAGRI）

ビッグデータの活用事例：産業分野別

2. 漁業

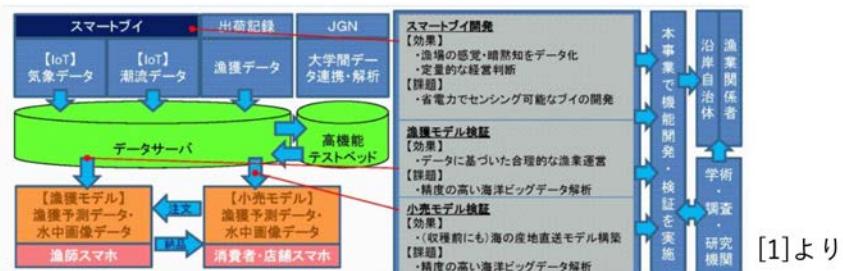
●定置網漁などに対するスマート漁業モデル事業[1]

気象データ、海流データ、漁獲データに基づく漁獲モデルの構築など

●遠洋かつお・まぐろ漁業のビッグデータ化[2]

海況予測や資源評価、魚群の来遊予測

●養殖魚管理、作業効率化[3]



[1]https://www.soumu.go.jp/main_sosiki/joho_tsusin/top/local_support/ict/jirei/2017_013.html

[2]https://www.jfa.maff.go.jp/j/kikaku/wpaper/h29_h/trend/1/t1_1_3_3.html

[3] <https://nissui.disclosure.site/ja/themes/144>

[1]-[3] Access 2021/3/7

東京大学 数理・情報教育研究センター 斎藤洋 2021 CC BY-NC-SA

31

- 総務省 | 海洋ビッグデータを活用したスマート漁業モデル事業 (2016年)
- 水産庁 | 遠洋かつお・まぐろ漁業でのICTを活用したビッグデータ化の取組 (2019年)
- ニッスイ | 養殖魚健康管理 - N-AHMS (2018年)

ビッグデータの活用事例：産業分野別

3. 製造業

●製造業のIoTプラットフォーム

ファンックの「Field System」

DMG 森精機の「ADAMOS」

Edgecross

●生産工程不良品検査、予測

[1]<https://www.automation-news.jp/2018/06/32869/>

[2]<https://www.omron.co.jp/technology/omrontechnics/2019/20190510-ogino.html>

[3]

https://www.hitachi.co.jp/products/it/lumada/usecase/case/lumada_uc_00016.html

[1]-[3] Access 2021/3/7

など

東京大学 数理・情報教育研究センター 斎藤洋 2021 CC BY-NC-SA

32

- オートメーション新聞 | 東芝デジタルソリューションズ 錠剤包装のさまざまな欠陥を検出する「PTP外観検査装置BLISPECTOR」 (2018年)
- オムロン | IoT技術によるシート検査装置から得られるデータの新たな現場活用手段 (2019年)
- 日立 | AIによる製品不良予兆診断 「Lumada」 ユースケース (2021年)

ビッグデータの活用事例：産業分野別

4. エネルギー産業

- 電力需要予測など発電事業への導入 [1][2]
- 電力メータ（スマートメータ）データを用いた他産業との連携[2]
 - 配送業の配達効率化
 - 家電利用状況管理
 - 銀行口座本人確認

[1] <https://eneken.ieej.or.jp/data/8680.pdf>

[2]

https://www.toshiba.co.jp/tech/review/2019/05/74_05pdf/a06.pdf

[3]https://www.enecho.meti.go.jp/about/special/johoteikyo/digitalization2019_3.html

[1]-[3]Access 2021/3/7

- 松原, 発電設備におけるAI&IoTの導入・開発状況と課題, IEEJ (2019年)
- 日本経済新聞 | 東芝 AIを活用した高精度な電力需要予測システムを開発 (2017年)
- 経済産業省 資源エネルギー庁 | 実績が見えてきた電力分野のデジタル化③～電力データ編 (2019年)

ビッグデータの活用事例：産業分野別

5. 医療

- 医学論文データに基づくガン種別の特定[1]
- 創薬分野[2]
- 画像認識による診断(ビッグデータの活用事例：画像データのページ参照)
- ヘルスケア支援[3]
- 血液データによるガン発見[4]
- ゲノムデータを用いた創薬、個別化医療

[1]<https://www.nikkei.com/article/DGXLZO05697850U6A800C1000000/>

[2] http://cbi-society.org/home/documents/seminar/2017to20/CBI391_Sawada.pdf

[3] <https://www.apple.com/jp/healthcare/>

[4] <https://grail.com/>

[1]-[4]Access 2021/3/7

- 日本経済新聞 | AI がん治療法助言 白血病のタイプ見抜く (2016年)
- 塩野義製薬 | 澤田, 製薬企業から見たAIの可能性 (2018年)
- Apple | ヘルスケアへの取り組み (カスタムApps, iPad, Apple Watch)
- GRAIL | 血液データによるがん発見ソリューション

4. 社会で活用されているデータ（種類と所有者）

社会で活用されているビッグデータの種類と所有者について解説する。データの種類として、具体的には調査データ、ログデータ、実験データ、観測データに大別される。またデータの所有(者)の概念として、独占禁止法、個人情報保護法、GDPR、データ発生元から二次加工、サードパーティ提供データの枠組み、ならびにデータに付随させるメタデータといった仕組みが存在する。

4.1 データの種類 [2]

4.1.1 調査データ [2]

調査データ (survey data) とは、アンケート調査のように調査を行う前に、その実施する目的を決め、その目的を達成するために収集されたデータを指す。アンケート調査には2つの大きな課題がある。1つは、いかにデータを得るのか、たとえばウェブ上か、対面あるいは郵送なのかである。もう1つは、誰からデータを得るのかという、対象の選択に関する課題である。

- [調査のチカラ | 調査データ検索](#)

後者は、統計学では最も大切な概念の1つの サンプリング (sampling) に直結した課題になる。サンプリングとは、対象となる集団の構成メンバーすべてからデータを得ることはできないため、データを得るメンバーを選別する方策のことをいう。前者の課題もサンプリングの問題と見なすこともできる。ITに対する理解度や利用環境の違いから生じる格差のことを [デジタルディバيد \(digital divide\)](#) と言い、調査時には [バイアス](#) がかかる注意しなければならない。

- [GMO research | サンプリングとは](#)
- [ELEMINIST | ICT化の裏に潜む「デジタルディバيد」とは 日本における情報格差の現状](#)
- [NEO MARKETING | リサーチの活用を妨げるバイアスとは？～確証バイアス～-ネオマーケティング-](#)

4.1.2 ログデータ [2]

ビッグデータの本格的利用は、検索エンジンの登場がきっかけである。情報収集の動きに拍車をかけたのが [eコマース \(Electric Commerce, 電子商取引\)](#) である。eコマース企業には、検索エンジン以上に、消費者が何に興味を持ち、どの商品との商品の相関が高そうかなど、ビジネス上きわめて価値のある情報が膨大に集まる。このように、ユーザや消費者の意識と関係なく計算機上に自動的に蓄積されるデータを [ログデータ \(logdata\)](#) と呼ぶ。本来の意味は、計算機が処理した作業の内容を、コンピュータの観点ですべて記録したデータを指す。したがって、ログデータの具体的な内容は以下に示すように非常に多岐にわたる。

- [J-Net 21 | EC（電子商取引）の基本戦略](#)
- [インテージテクノスフィア | 事例紹介 ログデータ活用の可能性](#)

(a) SNSの普及により、ログデータは個人情報とひもづけられ、ビッグデータの利用はさらに活発化した。個人のスマホのSNSアプリ(Apps)の画面には、当人の嗜好を読み取った広告画面が表示されている。移動体通信事業者は密に基地局を設置しているため、すべてのユーザの所在/移動記録が蓄積されている。この記録は、人の行動ログ (human activity log) になる。位置情報は個人情報そのものであるため、個人が特定されない形で、基地局ごとに何人いたかの情報が販売もされている。他方、多くの自動車にはナビゲーションシステムが搭載されており、自動車メーカーの集中管理システムが各車の位置情報を常時集めている。このような、走行する車に搭載されたさまざまなセンサのデータのことを [プローブデータ \(probe data\)](#) と呼ぶ。

- [愛知県ITS推進協議会 | プローブ情報](#)

(b) ログデータとして重要性を増しているのが、機械の稼働ログ (machine operation log) データであ

る。産業機械を主として、IoTセンサによって集められるデータがその代表的なものになる（事例：コマツ Komtrax）。企業ではログデータを解析することで、故障の前にトラブルの予告を行うサービスを行っている。また、故障の連絡があった際には作業員に、トラブルの原因の候補に対応した交換部品をリストアップするなど、ログデータの活用は作業の効率化に著しく貢献している。このように、機械の稼働ログのビッグデータの有効利用により、世界中で インダストリー4.0 を目指した競争が激しい。

- コマツ | 機械情報遠隔確認システム Komtrax
- デジタルトランスフォーメーションチャンネル | インダストリー4.0とは 製造業の未来を担う第
四次産業革命

4.1.3 実験データ [2]

研究の場面では、実験データ (experimental data) をよく取り扱う。実験データは、実験を目的としたセンサにより得られる情報のことである。近年の大きな特徴として、そのビッグデータ化を挙げる。

センサの空間解像度、時間分解能、エネルギー分解能などの劇的な向上により、データのサイズがそれまでの100倍、1,000倍になるケースも普通である。このため、それまでは専門家が目で判断していた作業では、データが生み出されるスピードにまったく追いつかず、データが死蔵されることも頻発している。

ICT の進歩により、これまで想像できなかった膨大で多様なデータが手に入り始めているが、人間の行動データについては、単に大量のデータが手に入っても、意味のある情報が得られにくい本質的な問題がある。その問題を解決しなければ、ビッグデータ研究は大きく発展しない可能性が高い（事例研究：寺澤2015）。

- 寺澤孝文, 教育ビッグデータの大きな可能性とアカデミズムに求められるもの-情報工学と社会科学のさらなる連携の重要性-, コンピュータ&エデュケーション, vol.38, p.28-38, 2015.

4.1.4 観測データ [2]

実験データのスケールアウトと同様のことは、観測データ (observation data) に対しても起こっている。

ビッグサイエンス (big science) と呼ばれる、総額数百億円から数千億円レベルの予算が必要な大型施設を用いた研究現場では、信じられないほどのビッグデータが産出されている。その多くは自然界における世界的新発見を目指したプロジェクトである（参考：CERN）。

- Spring-8/SACLA | CERN小史 -ビッグサイエンスの運命と技術と科学の相関

身近な観測データは、気象予報のニュースで示される人工衛星「ひまわり」の画像データ や、全国約1,300ヵ所に設置されたアメダス (AMeDAS) で計測された気温や降水量データ、また地震速報を可能にする2,000ヵ所程度の地震観測網による地震波形データである。

- NICT | NICTサイエンスクラウド ひまわり衛星プロジェクト
- 気象庁 | 過去の気象データ・ダウンロード
- 防災科学技術研究所 | 強震観測網(K-NET.KiK-net) 地震選択 & ダウンロード

4.2 データの所有者 [2]

4.2.1 独占禁止法、個人情報保護法、GDPR [2]

AIはビッグデータなしには成立しない。いかにビッグデータを獲得するかがビジネスの勝敗を決める。情報サービスにかかわるビジネスでは、ある業種の市場シェアトップ企業しか生き残れないとい

う「トップ総取り (Winner takes all)」の傾向が強い。ゆえに自社でデータを得られる（ユーザを囲い込める）システムを持つ企業が市場を独占する。SNS上で消費者やユーザが生み出すデータは、いったい誰のデータなのかという議論も活発である。

- [MarkeTRUNK | インターネットにおける「ウィナー・テイクス・オール」とは？](#)

消費者が不利にならないように、独占禁止法 (antitrust law) や個人情報保護法 (personal information protection law) による規制に各国政府は力を入れている。2018年に設定された [EUの一般データ保護規則 GDPR \(General Data Protection Regulation\)](#) も、この一連の問題に対応する形でEUが定めたデータ保護の枠組みである。

- [JETRO | EU 一般データ保護規則 \(GDPR\) について](#)

4.2.2 データ収集の枠組み（1次データ、2次データ、3次データ） [2]

データの所有者に関する意識の変化を受けて、法律、規則、社会的規範を守りながら、コストをかけずに効率的にデータを集められるかが、ビジネス上非常に大切になっている。データ収集の枠組みとして、データ発生元から順次加工され流通させる順番として、1次データ、2次データ、3次データの区別がある。

自社の業務や研究開発、また調査目的のために集めたデータのことを1次データ (primary data) と呼ぶ。貴重なデータであるためその管理には企業は細心の注意を払い、結果として、公開されないばかりか、競合他社が手に入れられることはほぼない。

2次データ (secondary data) とは外部のデータを指す。前項で解説したように、官公庁や国の研究機関、他社が保有している1次データなどのデータが相当する。気象庁が定期的に提供する、大規模なシミュレーションによる気象予報の数値データも2次データである。ウェザーニュースや日本気象協会のような気象予報を専門とする企業は、自社の1次データとこの2次データを融合して、自社独自の気象情報サービスを提供している。

- [データのじかん | 「1次データ」「2次データ」「3次データ」それぞれの違いと役割とは？](#)

3次データ (thirdparty data) とは、複数種類のデータを加工し、ユーザにとって使いやすくされたデータである。サードパーティーデータを解析することで、自社で収集した1次データからだけでは把握しきれなかった消費者の行動や関心の動向を把握するのに役立つ。サードパーティーデータを購入することは、2次データを購入し自社だけで解析するより、膨大なデータからどんな情報を集めたらよいのかという判断を下しやすすことや、煩雑なデータを整えるための労力を抑えられるというメリットがある（[引用サイト](#)）。

- [GMO research | マーケティングリサーチとは？基本用語と手法をわかりやすく解説](#)

4.2.3 メタデータ [2]

メタデータ (metadata) とは、データはいつ、だれが、いかにしてといったような、データが得られた状況を説明したデータである。

データの管理に用いられるさまざまな情報もメタデータと呼ばれる。たとえば実験データの場合、どの会社の測定装置を使ったか、実験したときの気温、湿度、気圧などの環境条件などがメタデータに相当する。メタデータは、データといっしょにデジタルデータとしてセットで保存しておくことが望ましい。データを得た際には面倒でも、さまざまなメタデータをデータに付属する（付随させる）、データのメタ化の習慣が大切である。

メタデータのような、データが得られた状況や環境に関する情報は、データの信頼性にかかわるため非常に重要である。研究メタデータ管理事例として [GakuNin RDM](#) がある。これは研究プロジェクト実施中に、個人の研究者あるいは研究グループが研究データや関連の資料を管理するための研究デ

ータ管理基盤である。既存のストレージや研究ソフトウェアと連携し、クローズドな空間で研究プロジェクトに関わるファイルのバージョン管理やメンバー内のアクセスコントロールが可能である。研究公正への対応としての研究証跡（真正性の確保、タイムスタンプ）を記録する機能や、ファイルを保存する機能を有する。

- [国立情報学研究所 | GakuNin RDM](#)（研究データ管理基盤）

5. 検索エンジンとSNS、オープンデータ [2]

検索エンジン（検索サービス）とSNS、オープンデータの活用事例について説明する。

5.1 検索エンジン [2]

ビッグデータの本格的な利用は、インターネット上に公開されたさまざまなサイトを上手にランキングする技術、つまり検索エンジン (search engine) により始まった。もはや私たちの生活に欠かせない道具となった検索エンジンにより、いくつかのキーワードを入力するだけでほしい情報を掲載しているサイトに効率よくたどり着ける。

個人に焦点をあてた広告は [ターゲティング広告 \(targeted advertising\)](#) と呼ばれている。eコマース (electronic commerce, 電子商取引) では、ウェブ上の個人向けの商品の提示、いわゆる [レコメンデーション \(recommendation\)](#) を最適化するために、消費者の購買履歴に関する膨大な量の情報収集を行っている。

- [Yahoo! JAPAN | ターゲティング広告とは？ターゲティングの種類や必要性を紹介](#)
- [NTTデータ数理システム | レコメンドシステムを最適化する技術とは？機械学習や数理最適化の活用事例](#)

ビッグデータから個人に適した情報サービスを提供する技術を、[パーソナライゼーション \(personalization\)](#) と呼ぶ。前述したレコメンデーションはその代表的な例である。

価値からビッグデータの流れは、消費者やユーザの嗜好性をデータにどのように変換し、ビッグデータとして蓄積するかの技術になる。たとえば、ウェブ上のサイトを訪問したユーザがどこに注目し、何をクリックし、そしてどのサイトを次に訪問したかの情報を効率よく収集する仕組みづくりがそうである。ECサイトによって収集・囲い込まれた年齢、性別、居住地、家族構成、職業などの個人情報と、サイト上のユーザの行動をコンピュータ上で統合する技術は [デモグラフィックデータ \(demographic data\)](#) と呼ばれる。

- [Adobe Marketo Engage | パーソナライゼーションとは？マーケティングに必要な理由と注意点](#)
- [ニュートラルワークス | QUERYY - デモグラ\(デモグラフィックデータ\)とは？サイコグラフィックとの違い](#)

5.2 SNS [2]

ソーシャルネットワーキングサービス SNS (social networking service) が広まって久しい。SNSにより、同じ嗜好性、共通の興味、同様の価値観を持ったグループが、通信履歴として特定される。またビッグデータの解析により、嗜好性、興味、価値観といった、抽象的概念を具体的に表す言葉やイメージが明らかになってくる。

SNS上のビッグデータの積極的活用により、意思決定の結果が多様でそして不確実な個人をターゲットにしたビジネスが可能になってきた。検索エンジンやSNSの利用により、膨大な数の細かいデモグラフィック情報まで、自動的かつコストをほとんどかけずにとれるようになってきた。消費者行動モデルの汎用性（モデルがどれくらい一般的に通用するかの）を追求しなくとも、個々の事象のケースバイケースの予測性能は、データを整理しパターン分類するだけでもかなり向上する。

このような、ビッグデータのパターン分類を基礎とする解決法を、[データ駆動型 \(data driven\) アプローチ](#)という。標準的なモデルを先に構築し、状況に応じてそのモデルを変化させる（演繹的）ではなく、まず豊富に用意された行動の変容に関する条件とその結果のペアをビッグデータから大量に得たあと、それらをマイニング・整理することで対応関係を作成し、その対応関係から未来の行動を予測する（帰納的）。

- [ZIP Forecasting | データ駆動型アプローチ：ビジネスへの定義とプロセス](#)

5.3 オープンデータ活用事例 [5]

企業によるデータの集め方に関してさまざまな規制が強化されつつある中、国や研究機関が集めたデータを多くの人に使ってもらおうという動きが活発である。このようなデータをオープンデータ (opendata) という。特に、国や地方自治体の集めたデータは、個人情報に十分に留意しながら加工して公開すれば、住民の日常生活に役立つばかりか、企業は新しいビジネスを考えられる。また、税金で成り立っている研究資金によって得られた実験データや観測データは、公開を原則とする流れも強まっている。国費を投じて得られる調査データに関しても、法律や諸々の社会的制約の中で、広く使ってもらう努力がなされている。

- [オープンデータ活用事例28選とおすすめデータセットまとめ（宇宙） 《利用条件》](#)

1.オープンデータを基盤に新規サービス創出(新価値創造型)

- [新型コロナダッシュボード](#)（国内の新型コロナウイルスの感染者数、病床数と使用率を都道府県ごとに一覧で確認）
- [Beyond Floods](#)（アメリカ自分が住んでいる場所の洪水リスクを把握）
- [Viomedo](#)（ドイツ臨床試験と臨床試験の条件に合う患者とのマッチングプラットフォーム(DE)）
- [Spotify](#)（音楽に関するメタデータ）
- [MUSICBRAINZ](#)（音楽作品のメタデータ投稿と公開）
- [Fooducate](#)（米国労働統計局、食品栄養サービス 栄養価を評価）
- [カーリル](#)（国内各自治体の持つ図書館蔵書データベースを束ねた横断検索サービス）
- [Young Europeans](#)（ヨーロッパの若者がさまざまな指標でEUの平均的な若者と比較できるサービス）
- [GRASP EARTH](#)（全球変化検出サービス - 衛星データを用いて任意の場所の地理空間情報を時系列で比較）

2.既存サービスにオープンデータをプラス(ビジネスアップデート型)

- [Coaido 119](#)（119番通報をしながら現在地周辺の必要な施設、人にSOSを発信できる緊急情報共有アプリ）
- [駅前不動産](#)（物件探しサービスに校区情報のオープンデータを利用。今見ている物件がどの公立学校の校区内なのか確認できる）
- [NeighborhoodScout](#)（アメリカ犯罪発生率、住宅価格、学校の質などのデータを可視化）
- [スマイティ](#)（株式会社カカクコムが提供するスマイティの「住みやすい街」）
- [komoot](#)（アウトドアを楽しむ人のため、その人に合わせたルート検索とナビゲーションを行うアプリ。アウトドアスポット情報、人工衛星から取得した標高データ、Open Street Mapを利用）
- [ruprun](#)（ユーザーの希望に沿ったランニングコースを提案するアプリ。銭湯の場所や観光施設、公衆トイレなど、主に公共施設のオープンデータを使用）
- [いこーよ](#)（おでかけスポットやイベントの検索と、おでかけスポットの口コミ投稿を共有。自治体が提供するオープンデータを利用）
- [Zaim](#)（家計簿・会計アプリ。国や各自治体が提供する給付金・手当・控除情報を利用。）

- [LIVE JAPAN PERFECT GUIDE](#)（文化財一覧、観光施設一覧、公衆無線LANアクセスポイント一覧、公衆トイレ一覧、AED設置箇所一覧などのオープンデータを利用）

3.地域観光・地域行政を変えるオープンデータ活用事例

- [chariP naVi](#)（京都の歴史や文化、伝統的な美しい街並みを自転車で楽しむための京都案内サービス。駐輪場一覧、京都市認定レンタサイクル店一覧、観光施設情報のオープンデータを利用。）
- [Bmaps](#)（文化施設におけるエレベーター、多目的トイレのバリアフリー情報のオープンデータを用いたバリアフリー地図アプリ）
- [Wheelog](#)（車いすの利用者同士が、GPSを活用して車いすで移動したルートを地図に記録し、車いすが通行できる道を共有できるアプリ）
- [My City Report](#)（利用者が地域の公共インフラに関する問題を投稿すると該当地域の担当課に連絡が届き、対応の実施有無をアプリ内で確認ができるサービス）
- [バスロケ](#)（停留所情報（名称・緯度経度）、運行系統、系統番号、現在位置、遅延情報、行先情報（リアルタイム情報）などのデータを用いて、事務所での正確な運行管理を実現）
- [除雪車ナビ](#)（市道除雪路線データ、除雪車走行データを用いて除雪車が今どこにいるかを把握できるサービス）
- [セーフティマップ](#)（HONDAが取得する急ブレーキの発生箇所のデータ、自治体の交通事故情報等のデータを用いて事故多発箇所や要注意箇所を地図上に提示するサービス）
- [交通事故予測アプリ](#)（NTT社用車のドラレコデータ、香川県警の交通事故データ、他のオープンデータを用いて交通事故を予測）
- [全国避難所ガイド](#)（全国の避難所、広域避難場所、一時避難場所、帰宅困難者一時滞在施設、津波避難施設等に関する情報のオープンデータを集約）
- [Intelligent Zoning Engine](#)（ベルリン 小学校区の設定問題に対応するために作られたサービス）

memo