

数理・データサイエンス・AI  
モデルカリキュラム準拠の教材について

寒野 善博

東京大学 数理・情報教育研究センター

数理・データサイエンス・AI 教育強化拠点コンソーシアム  
2024 年度 関東ブロック対面ワークショップ

2024 年 10 月 7 日

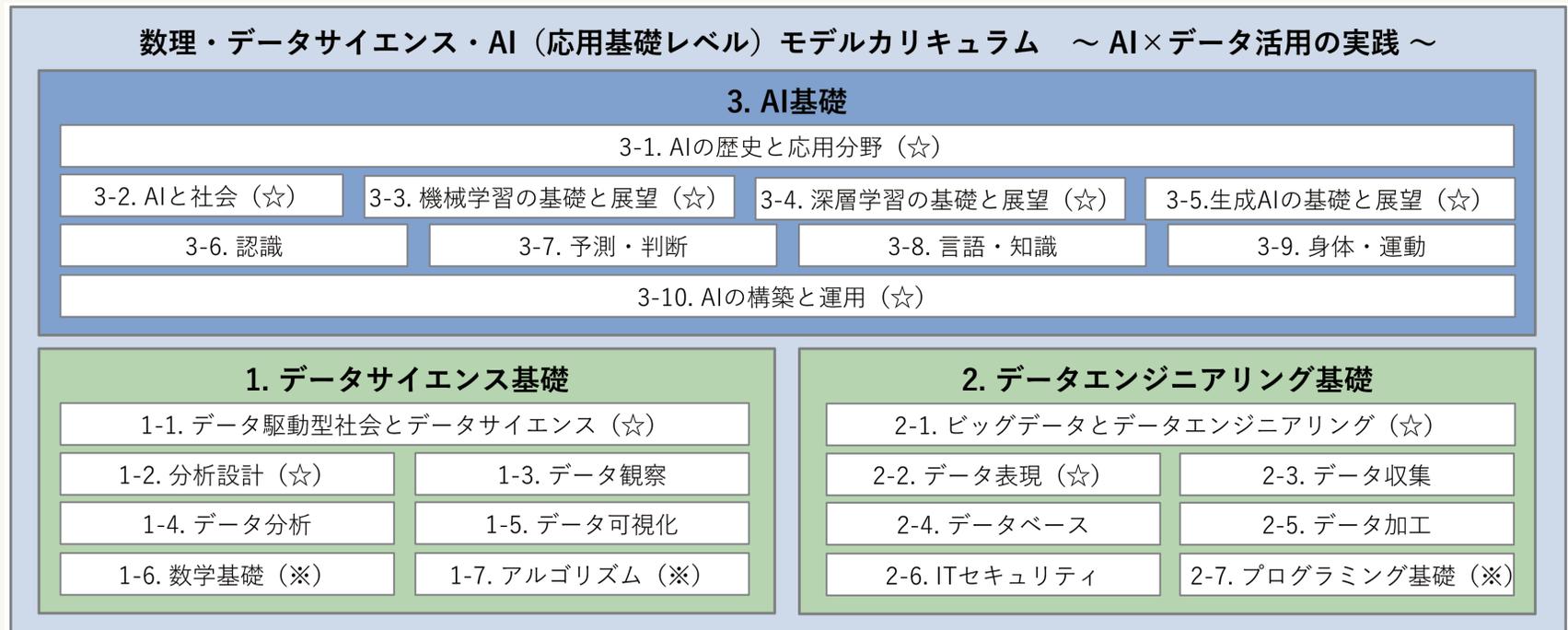
# モデルカリキュラム

- 「数理・データサイエンス・AI教育強化拠点コンソーシアム」が制定
  - リテラシーレベル（2020年4月；2024年2月に改訂）
  - 応用基礎レベル（2021年3月；2024年2月に改訂）
- リテラシーレベルの構成

導入	<b>1. 社会におけるデータ・AI利活用</b>	
	1-1. 社会で起きている変化	1-2. 社会で活用されているデータ
	1-3. データ・AIの活用領域	1-4. データ・AI利活用のための技術
	1-5. データ・AI利活用の現場	1-6. データ・AI利活用の最新動向
基礎	<b>2. データリテラシー</b>	
	2-1. データを読む	2-2. データを説明する
	2-3. データを扱う	
心得	<b>3. データ・AI利活用における留意事項</b>	
	3-1. データ・AIを扱う上での留意事項	3-2. データを守る上での留意事項
選択	<b>4. オプション</b>	
	4-1. 統計および数理基礎	4-2. アルゴリズム基礎
	4-3. データ構造とプログラミング基礎	4-4. 時系列データ解析
	4-5. テキスト解析	4-6. 画像解析
	4-7. データハンドリング	4-8. データ活用実践（教師あり学習）
	4-9. データ活用実践（教師なし学習）	

# モデルカリキュラム

- 「数理・データサイエンス・AI 教育強化拠点コンソーシアム」が制定
  - リテラシーレベル（2020年4月；2024年2月に改訂）
  - 応用基礎レベル（2021年3月；2024年2月に改訂）
- 応用基礎レベルの構成



## モデルカリキュラムの内容：リテラシーレベルの場合

節のタイトル	キーワード（知識・スキル）
1-1. 社会で起きている変化	<ul style="list-style-type: none"> <li>・ビッグデータ、IoT、AI、生成AI、ロボット</li> <li>・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</li> <li>・第4次産業革命、Society 5.0、データ駆動型社会</li> <li>・複数技術を組み合わせたAIサービス</li> <li>・人間の知的活動とAIの関係性</li> <li>・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</li> </ul>
1-2. 社会で活用されているデータ	<ul style="list-style-type: none"> <li>・調査データ、実験データ.....</li> </ul>
<ul style="list-style-type: none"> <li>・</li> <li>・</li> </ul>	<ul style="list-style-type: none"> <li>・</li> <li>・</li> </ul>
4-9. データ活用実践（教師なし学習）	<ul style="list-style-type: none"> <li>・教師なし学習によるグルーピング.....</li> </ul>

教材：全般的な説明

# 教材：数理・情報教育研究センター（東大）で開発

- 公開場所：[http://www.mi.u-tokyo.ac.jp/6university\\_consortium.html](http://www.mi.u-tokyo.ac.jp/6university_consortium.html)
- 「センターのwebページ」 → 「大学間コンソーシアムのページ」

数理・情報教育研究センター  
Mathematics and Informatics Center

東京大学  
THE UNIVERSITY OF TOKYO

トップ  
ENGLISH

センター概要 | メンバー | **大学間コンソーシアム** | 産業界コンソーシアム | 関連講義 | e-learning教材 | 社会人教育 | DSコモンズ

## 数理・データサイエンス・AI教育強化拠点コンソーシアム

MIセンターは、2022年度政府予算に盛り込まれた「数理・データサイエンス・AI教育の全国展開の推進」事業の東京大学における実施主体です。同事業で選定された29大学（拠点校11大学、特定分野校18大学）のコンソーシアムの幹事校として、大学、産業界、研究機関等と幅広くネットワークを形成し、地域や分野における先進的教育モデルの拠点として、数理・データサイエンス・AIの実践的教育の全国普及に努めます。同時に、この分野を牽引できる国際競争力のある人材および産学で活躍できるトップクラスのエキスパート人材の育成を目指します。

[\[コンソーシアムホームページ\]](#)

### 数理・データサイエンス・AIの活用事例動画

本動画集は数理・データサイエンス・AIリテラシーレベル教材の導入となるような活用事例を収集したものです。数理・データサイエンス・AIリテラシーレベル教材との対応は下の教材一覧をご覧ください。

[人工知能 \(AI\) の活用事例動画一覧](#)

**数理・データサイエンス・AI教材**

# 教材：レベルごとに整理

## ● リテラシーレベル

### リテラシーレベル教材

章	節	教材	実習用補助教材
1. 社会におけるデータ・AI活用	1-1. 社会で起きている変化	<a href="#">スライド</a>	
	1-2. 社会で活用されているデータ	<a href="#">スライド</a>	
	1-3. データ・AIの活用領域	<a href="#">スライド</a>	<a href="#">人工知能（AI）の活用事例動画</a>
	1-4. データ・AI利活用のための技術	<a href="#">スライド</a>	
	1-5. データ・AI利活用の現場	<a href="#">スライド</a>	

## ● 応用基礎レベル

### 応用基礎レベル教材

章	節	教材	実習用補助教材
1. データサイエンス基礎	1-1. データ駆動型社会とデータサイエンス	<a href="#">スライド</a>	
	1-2. 分析設計	<a href="#">スライド</a>	
	1-3. データ観察	<a href="#">スライド</a>	
	1-4. データ分析	<a href="#">スライド</a>	<a href="#">補助資料</a>
	1-5. データ可視化	<a href="#">スライド</a>	<a href="#">補助教材</a>

# 教材：内容

1-4. データ・AI利活用のための技術	<a href="#">スライド</a>	<a href="#">人工知能（AI）の活用事例動画</a>
1-5. データ・AI利活用の現場	<a href="#">スライド</a>	
4-4. 時系列データの解析	<a href="#">スライド</a>	<a href="#">データ</a> , 解析例 ( <a href="#">pdf</a> , <a href="#">excel</a> )
4-5. テキスト解析	<a href="#">スライド</a>	
4-6. 画像解析	<a href="#">スライド</a>	Pythonサンプルコード ( <a href="#">pdf1</a> , <a href="#">pdf2</a> , <a href="#">pdf3</a> )
4-7. データハンドリング	<a href="#">スライド</a>	
4-8. データ活用実践（教師あり学習）	<a href="#">スライド</a>	<a href="#">データ解析例</a>
4-9. データ活用実践（教師なし学習）	<a href="#">スライド</a>	<a href="#">データ解析例</a>

- 講義スライド（このあと ご紹介）
- 活用事例の動画
- データ解析例（Python や R による例）
- Python サンプルコード

# 講義スライドの構成と特徴

# 講義スライド形式のPDFファイル

章	節	教材	実習用補助教材
1. 社会におけるデータ・AI利活用	1-1. 社会で起きている変化	<u>スライド</u>	
	1-2. 社会で活用されているデータ	<u>スライド</u>	
	1-3. データ・AIの活用領域	<u>スライド</u>	
	1-4. データ・AI利活用のための技術	<u>スライド</u>	

- 表紙

## 1-1 社会で起きている変化

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

- データサイエンスやAIとはどういう分野なのでしょうか？なぜ社会の関心を集めているのでしょうか？
- 本節では社会におけるデータ・AIの利活用例を幅広く学ぶことで、社会で起きている変化を知り、データサイエンスやAIを学ぶことの意義を理解することを目標とします
- 特にAIを活用した新しいビジネスやサービスは、複数の技術が組み合わさって実現していることに注目して下さい

## 本教材の目次

1. データサイエンス入門	4
2. 応用分野（自然科学）	13
3. 応用分野（社会科学）	17
4. 応用分野（ビジネス）	24
5. データサイエンティストとAIの関係	35
6. 参考文献	50

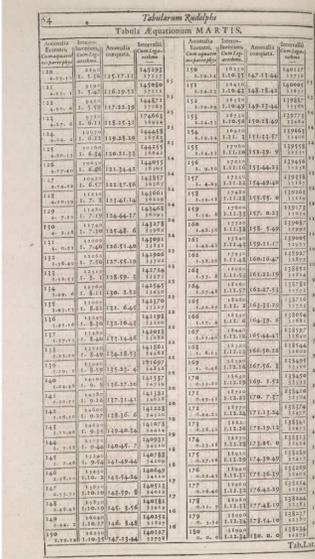
## 1-1-1 データサイエンス入門

## データサイエンスとは

- データを有効活用し、数理モデリングや計算技術と適用ドメインの専門知識を結合することで新たな知識を生み出し、その活用のシナリオを導き出すことです
- なぜ注目を集めているのでしょうか？

## 古くからあります

- 例えば17世紀の天文学者であるヨハネス・ケプラーは
- (1) データの有効活用：ティコ・ブラーエが長年観測した惑星の運動データを用い 
- (2) 数理モデル・計算技術・適用ドメインの専門知識：天文学の知識を活用し、手計算で惑星運動を分析し
  - 計算機はおろか消しゴムすらまだありません
- (3) 活用のシナリオ：仮説検証や知識発見などを通じてその後の天文学に大きな影響を与えました
  - ケプラーの法則はニュートンの万有引力へと繋がっていきます
- なぜ今注目を集めているのでしょうか？



The image shows a page from Johannes Kepler's 'Rudolphine Tables' (1627), specifically the 'Tabula Aequationum MARTIS' (Table of Equations for Mars). It is a dense table of numerical data, likely representing the positions and motions of Mars as calculated by Kepler. The table is organized into columns with various headings in Latin, and the data is presented in a grid-like format with many rows of numbers.

出典：  
[Kepler1627]  
<https://www.loc.gov/item/49038330/>

東京大学 数理・情報教育研究センター 久野遼平 斎藤洋 近藤亮磨 2024 CC BY-NC-SA

6

- … このような構成の，スライド教材です。

## 教材の特徴 (1/2)

- モデルカリキュラムに完全準拠
  - **すべての**キーワード（知識・スキル）を解説

### キーワードとの対応表

キーワード	章-節	ページ
1次データ	1-2	12
1変数関数の微分と積分	4-1	55-90
2軸グラフ	1-4	7
2次データ	1-2	12
AI	1-1	33-34
AIサービスの責任論	3-1	31
AI最新技術の活用例	1-6	5, 6, 7, 8, 9, 10
AI社会原則	3-1	30
AI等を活用した新しいビジネスモデル	1-6	5, 6, 7, 8, 10, 11
AIとビッグデータ	1-4	22

## 教材の特徴 (2/2)

- 教育目的ならば、ページ単位で自由に利用可能
  - CC ライセンス「表示」「非営利」「継承」
- 適宜、改善や修正も...

### リテラシーレベル教材についてのアンケート

リテラシーレベル教材の利用につきまして、今後の参考のため、アンケートにご回答ください。メールアドレス（任意）をご記入いただいた大学には、教材の更新情報などをご連絡させていただきます。予定です。

利用の状況や、ご意見をお気軽にご記入ください。後日、追加のご意見やご要望も歓迎いたします。

### 応用基礎レベル教材についてのアンケート

「モデルカリキュラム改訂」に関連して

# モデルカリキュラム（リテラシーレベル，応用基礎）が今年2月に改訂

- 改訂の背景
  - プログラミング教育：小・中学校
  - 「情報I」必修化：高等学校
  - 生成AIなど：社会動向
- 改訂の内容（下記の他にも，キーワードの調整・追加あり）
  - 「情報I」とのキーワードの重複：既習を理由に削除はしない
  - 生成AI
    - リテラシーレベル：基本の考え方・仕組みに関するキーワードを追加
    - 応用基礎：「生成AIの基礎と展望」を新たなコア学修項目に

委員に公私立大学、産業界、生成AI、高等学校の関係者を参画させるとともに、専門家からの意見聴取の機会を設け、幅広い合意形成に努め、この度、以下のとおり「数理・データサイエンス・AI(リテラシーレベル)モデルカリキュラム～ データ思考の涵養 ～」の改訂を行いました。

[数理・データサイエンス・AI\(リテラシーレベル\)モデルカリキュラム ～ データ思考の涵養 ～ \(2024年2月22日改訂\)※改訂箇所明示版](#)

[数理・データサイエンス・AI\(リテラシーレベル\)モデルカリキュラム ～ データ思考の涵養 ～ \(2024年2月22日改訂\)※クリーン版](#)

[http://www.mi.u-tokyo.ac.jp/consortium/model\\_literacy.html](http://www.mi.u-tokyo.ac.jp/consortium/model_literacy.html)

# 教材（講義スライド）では、「リテラシーレベル」の改訂に対応

## ● キーワード“生成 AI”が追加された例

節のタイトル	キーワード（知識・スキル）
1-1. 社会で起きている変化	<ul style="list-style-type: none"><li>・ ビッグデータ、IoT、AI、“生成 AI”、ロボット</li><li>・ 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</li><li>・ 第4次産業革命、Society 5.0、データ駆動型社会</li><li>・ …</li></ul>

4-7. データハンドリング	<ul style="list-style-type: none"><li>・ …</li><li>・ データの抽出</li><li>・ データの結合</li><li>・ 名寄せ</li><li>・ “生成 AI を活用したデータ加工”</li></ul>
----------------	--

# “生成 AI” の追加への対応例

- 「1-1. 社会で起きている変化」への追加
- 教材の改訂
  - 対応するスライドを新たに作成しました：

## 生成AI

- インターネットスケールで手に入るデータはテキストだけではありません
- 例えば画像とキャプションのデータを用いて「アボガドのような椅子」や「ヤマアラシのような立方体」など複雑なテキスト入力に対して画像を生成することができるようになっていきます
  - OpenAIのDallE
    - <https://openai.com/blog/dall-e/>
  - 画像生成は一般用のPCでも実行できるようになっています
    - Stable Diffusion
      - <https://ja.stability.ai/stable-diffusion>

# 「生成 AI を活用したデータ加工」の解説 (1/4)

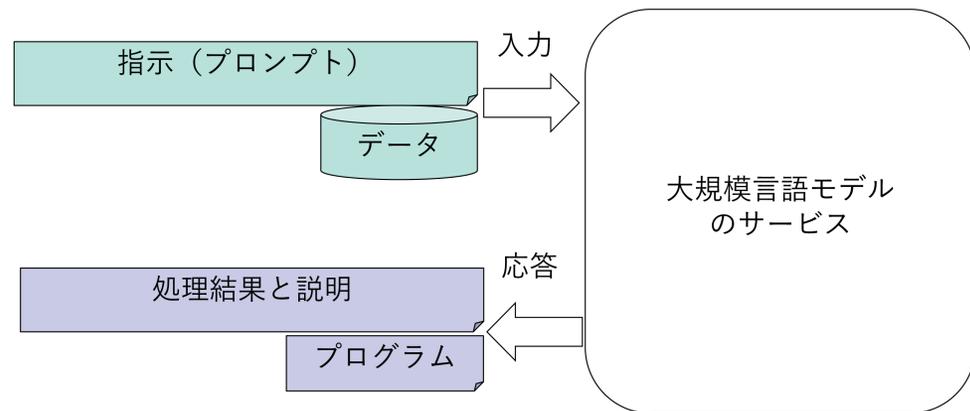
## 生成AIを利用したデータ加工

- 主に基盤モデルに基づく生成モデルを用いてテキスト、画像、その他のコンテンツを生成する人工知能 (AI) の技術を生成AIと呼びます。
  - 基盤モデルは、大規模なデータをもとに事前学習された機械学習のモデルです。
  - 例えば大規模言語モデル (Large Language Model, LLM) は、大規模なテキストデータをもとに学習された基盤モデルです。
  - 大規模言語モデルは、与えられた文脈 (単語の接続) から次の単語を予測するという事前学習により学習されます。
  - 加えて、指示に対して適切な応答を返すことや、倫理的・社会的に適切な応答を返すことを学習させることで、大規模言語モデルはさまざまなタスク (問題) を解くことができます。
- 大規模言語モデルの応用として、プログラミングに  
たデータの処理があります。

東京大学 数理・情報教育研究センター 森純一郎 2020 CC BY-NC-S

## 生成AIを利用したデータ加工

- 大規模言語モデルに適切な指示 (プロンプト) とデータを与えることで、モデルが生成したプログラムをもとにデータの集計・検索・抽出や前処理の作業を対話的に行うことができます。



東京大学 数理・情報教育研究センター 森純一郎 2020 CC BY-NC-SA

# 「生成 AI を活用したデータ加工」の解説 (2/4)

## 生成AIを利用したデータ加工

大規模言語モデルのサービスとの対話の例

- 添付のファイルを読み込んでください  
(教育用標準データセット\*の基本素材 (SSDSE-E) のCSVファイルをアップロード)

\* <https://www.nstac.go.jp/use/literacy/ssdse/>

- CSVファイルを正常に読み込むことができました。ファイルは日本の各種統計データを含んでおり、都道府県ごとの人口、出生率、医療施設数、消費支出など多岐にわたる情報が含まれています。具体的なデータ処理や分析が必要な場合は、どのように進めたいか教えてください。
- 総人口の列について北海道から沖縄県までの47都道してください
- 47都道府県の総人口の平均値は約2,658,426人です

東京大学 数理・情報教育研究センター 森純一郎 2020 CC BY-NC-S

## 生成AIを利用したデータ加工

大規模言語モデルのサービスとの対話の例 (つづき)

- この平均値以上の総人口の都道府県を抽出してください
- 47都道府県の平均値以上の総人口を持つ都道府県は以下の通りです：
  - 北海道：5,140,000人
  - 茨城県：2,840,000人
  - 埼玉県：7,337,000人
  - 千葉県：6,266,000人
  - 東京都：14,038,000人
  - 神奈川県：9,232,000人
  - 静岡県：3,582,000人
  - 愛知県：7,495,000人
  - 大阪府：8,782,000人
  - 兵庫県：5,402,000人
  - 広島県：2,760,000人
  - 福岡県：5,116,000人

東京大学 数理・情報教育研究センター 森純一郎 2020 CC BY-NC-SA

25

# 「生成 AI を活用したデータ加工」の解説 (3/4)

## 生成AIを利用したデータ加工

大規模言語モデルのサービスとの対話の例 (つづき)

- 47都道府県の総人口の第1四分位点, 第3四分位点, 四分位範囲を求めてください
- 47都道府県の総人口について, 以下の統計値が求められました:
  - 第1四分位点 (Q1) : 1,046,500人
  - 第3四分位点 (Q3) : 2,655,000人
  - 四分位範囲 (IQR) : 1,608,500人

## 生成AIを利用したデータ加工

大規模言語モデルのサービスとの対話の例 (つづき)

- これらの値から外れ値の基準をつくれますか

- はい, これらの四分位数を使って外れ値の基準を作成することができます. 外れ値の基準は, 以下のように計算されます:
  - 下限の外れ値の基準:  $Q1 - 1.5 \times IQR$
  - 上限の外れ値の基準:  $Q3 + 1.5 \times IQR$
- これらの値を計算してみましょう. 47都道府県の総人口に基づいた外れ値の基準は次のようになります:
  - 下限の外れ値の基準: -1,366,250人 (実際の応用では, 人口が負の値になることはありませんので, この下限は実用的ではありません)
  - 上限の外れ値の基準: 5,067,750人 この上限値を超える人口を持つ都道府県は外れ値と考えられます. この基準により, 特に人口の多い都道府県が識別されます.

東京大学 数理・情報教育研究センター 森純一郎 2020 CC BY-NC-S

東京大学 数理・情報教育研究センター 森純一郎 2020 CC BY-NC-SA

27

# 「生成 AI を活用したデータ加工」の解説 (4/4)

## 生成AIを利用したデータ加工

大規模言語モデルのサービスとの対話の例 (つづき)

- 具体的にどの都道府県の値が外れ値ですか

- 人口が外れ値として識別された都道府県は以下の通りです：

- 北海道：5,140,000人
- 埼玉県：7,337,000人
- 千葉県：6,266,000人
- 東京都：14,038,000人
- 神奈川県：9,232,000人
- 愛知県：7,495,000人
- 大阪府：8,782,000人
- 兵庫県：5,402,000人
- 福岡県：5,116,000人

これらの都道府県は、47都道府県の総人口の上限（5,067,750人）を超える人口を持っています。これが集中している地域であり、日本の人口分布の特

## 生成AIを利用したデータ加工

- 生成AIとして大規模言語モデルを利用したこれらの処理は、実際はモデルが自ら生成したプログラムによって行われています。
- 大規模言語モデルには、実際は存在しない情報を生成してしまうハルシネーション（幻覚）という問題が存在します。
- 大規模言語モデルが生成したプログラムとその処理結果についても、手続きや結果に誤りがないか十分に留意して利用する必要があります。
- また、生成AIのサービス提供者が入力した情報をどのように取り扱っているのかを十分に確認した上で利用する必要があります。
  - 特に、機密情報や個人情報の扱いには十分に留意が必要です。

## その他の追加キーワードの例 (1/2)

- 例として、「2-3」では“赤字”が新たに追加に.

節のタイトル	キーワード（知識・スキル）
2-3. データを扱う	<ul style="list-style-type: none"><li>▪ “データの取得（機械判読可能なデータの作成・表記方法）”</li><li>▪ データの集計（和、平均）</li><li>▪ 第4次産業革命、Society 5.0、データ駆動型社会</li><li>▪ データの並び替え、ランキング</li><li>▪ データ解析ツール（スプレッドシート、“BI ツール”）</li><li>▪ 表形式のデータ（csv）</li></ul>

## その他の追加キーワードの例 (2/2)

- 例として、「2-3」では“赤字”が新たに追加に.

節のタイトル	キーワード（知識・スキル）
2-3. データを扱う	・“データの取得（機械判読可能なデータの作成・表記方法）”

- 教材での対応

- 解説スライドを追加しました：

### データの取得（機械判読可能なデータ）

- ・ データ分析を行うためのデータ収集では、データの正確性に注意するとともに、機械判読可能な形式のデータを取得する必要があります。
- ・ 機械判読可能なデータとは、コンピュータが直接解析や処理が可能な形式で提供されているデータのことを指します。このようなデータは、人間が読むための文書とは異なり、データ分析やAIの学習など、さまざまな自動処理に直接利用することができます。
- ・ 機械判読可能なデータ形式の代表的なものとして、後述のcsv形式のデータが挙げられます。

その他の講義スライドの特徴

# 図や画像を用いて 具体的にわかりやすく解説 (1/2)

## カラー画像をHSV空間に変換

original image



H image



時計台は赤

S image



木々は時計台より鮮やか

明るい方が鮮やか

V image



白黒写真として使われる

東京大学 数理・情報教育研究センター 山肩洋子 2020 CC BY-NC-SA

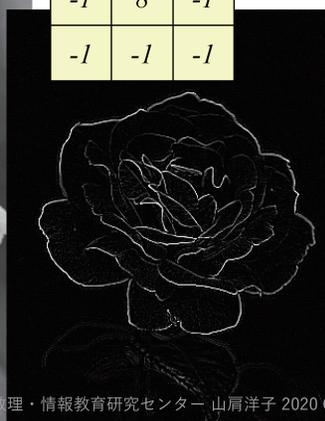
## エッジ抽出：ラプラシアンフィルタ

その画素の値が周辺の画素と違うほど大きな値とすることによって、周辺との輝度が大きく変化する輪郭線を取り出す

3x3 ラプラシアンフィルタ

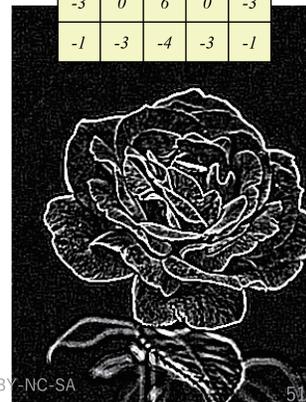
-1	-1	-1
-1	8	-1
-1	-1	-1

入力画像(320x480)



5x5 ラプラシアンフィルタ

-1	-3	-4	-3	-1
-3	0	6	0	-3
-4	6	20	6	-4
-3	0	6	0	-3
-1	-3	-4	-3	-1

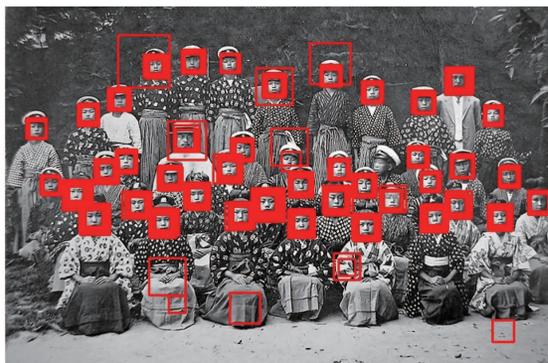


東京大学 数理・情報教育研究センター 山肩洋子 2020 CC BY-NC-SA

51

# 図や画像を用いて 具体的にわかりやすく解説 (2/2)

## 顔領域の検出結果



- 画像サイズを徐々に小さくしながらモデルと適合するかをスキャン
- 顔ではないところも誤検出される
- 真の顔領域であれば、画像のサイズを変えるたびに、顔領域として何度も検出される



2回以上検出された領域を顔として検出

ref. (2020/4/6): Wikimedia commons: File:Kasahara Saitama Kasahara Jinjo Elementary School 1920 1.jpg パブリックドメイン  
[https://ja.wikipedia.org/wiki/%E3%83%95%E3%82%A1%E3%82%A4%E3%83%AB-Kasahara\\_Saitama\\_Kasahara\\_Jinjo\\_Elementary\\_School\\_1920\\_1.jpg](https://ja.wikipedia.org/wiki/%E3%83%95%E3%82%A1%E3%82%A4%E3%83%AB-Kasahara_Saitama_Kasahara_Jinjo_Elementary_School_1920_1.jpg)



東京大学 数理・情報教育研究センター 山肩洋子 2020 CC BY

## 画像認識の例：手書き文字認識

- 画像認識初期のタスク
- よく使われるデータはMNIST (Mixed National Institute of Standards and Technology database)
  - 「0~9」の10種類の数字の認識 = 10クラス分類タスク
  - 各画像に数字が1つ記入
  - 解像度は20x20 pixel
  - 訓練データ：60,000枚
  - 評価データ：10,000枚
- 間違いやすいサンプルも含む

4 3 8 2  
4→6 3→5 8→2 2→1  
4 8 7 5  
9→4 8→0 7→8 5→3

3 6 8 1 7 9 6 6 9 1  
6 7 5 7 8 6 3 4 8 5  
2 1 7 9 7 1 2 8 4 5  
4 8 1 9 0 1 8 8 9 4  
7 6 1 8 6 4 1 5 6 0  
7 5 9 2 6 5 8 1 9 7  
2 2 2 2 2 3 4 4 8 0  
0 2 3 8 0 7 3 8 5 7  
0 1 4 6 4 6 0 2 4 3  
7 1 2 8 1 6 9 8 6 1

ref. (2020/4/6): THE MNIST DATABASE of handwritten digits <http://yann.lecun.com/exdb/mnist/>  
Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. "Gradient-based learning applied to document recognition." Proceedings of the IEEE, 86(11):2278-2324, November 1998.

東京大学 数理・情報教育研究センター 山肩洋子 2020 CC BY-NC-SA

66

# 数式を用いない説明, スプレッドシートでの計算法なども解説

## 相関行列

- データの全ての項目に対して、任意の2種類の相関をマトリックスで表示したものを相関行列といいます。

	緯度	経度	地震の深さ (km)	マグニチュード	計測地点数
緯度	1.00	-0.36	0.03	-0.05	-0.00
経度	-0.36	1.00	0.14	-0.17	-0.05
地震の深さ (km)	0.03	0.14	1.00	-0.23	-0.07
マグニチュード	-0.05	-0.17	-0.23	1.00	0
計測地点数	-0.00	-0.05	-0.07	0	1

「経度」と「地震の深さ」の相関が0.14ということになります。

東京大学 数理・情報教育研究センター 荻原哲平 2020 CC BY-

同じ項相関は1になります。

## データに対する操作 (和、平均)

### 和をとる

- 和をとる範囲を指定してsum関数を用いて計算します。
- 例えば、右のデータで7月の日射量の合計を計算するときは、適当なセル上で  
=sum(D2:D32)  
等と打つと和を計算します。

< 気候データ >

fx	A	B	C	D	E	F
1	月	日	気温	日射量	湿度	
2	7	1	28	14.08	72	
3	7	2	28.7	14.54	73	
4	7	3	28.4	13.18	72	
5	7	4	27.5	7.82	75	
6	7	5	26.7	3.49	80	
7	7	6	21.6	1.42	103	
8	7	7	24.9	8.11	81	
9	7	8	27.3	10.48	77	
10	7	9	27.2	9.08	89	

=sum(D2:D32)  
と打つと日射量の合計が計算される

E	F	G	H
湿度			
72		306.25	
73			
72			
...			

### 平均をとる

- 平均をとる範囲を指定してaverage関数で計算します。
- 右のデータで7月の平均気温を計算するには、  
=average(C2:C32)  
と打ちます。

< 気候データ >

fx	A	B	C	D	E	F
1	月	日	気温	日射量	湿度	
2	7	1	28	14.08	72	
3	7	2	28.7	14.54	73	
4	7	3	28.4	13.18	72	
5	7	4	27.5	7.82	75	
6	7	5	26.7	3.49	80	
7	7	6	21.6	1.42	103	
8	7	7	24.9	8.11	81	
9	7	8	27.3	10.48	77	
10	7	9	27.2	9.08	89	

=average(C2:C32)  
と打つと気温の平均が計算される

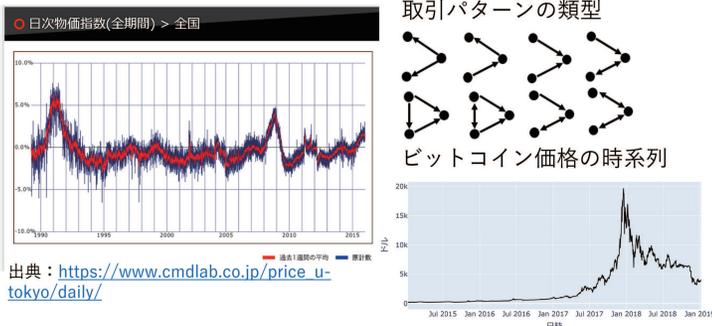
E	F	G	H
湿度			
72		306.25	
73			
72		28.21	
75			
80			
103			

東京大学 数理・情報教育研究センター 荻原哲平 2020 CC BY-NC-SA

# 幅広い分野の事例を用いて解説

## 経済学

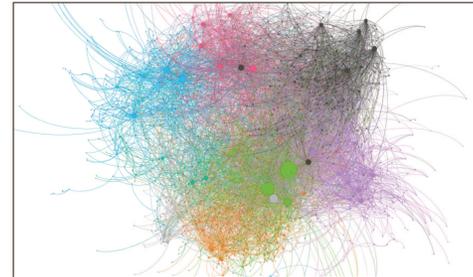
- 経済学でもビッグデータを活用した研究が行われています
- POSデータを用い物価をリアルタイムで計測する研究者もいます[渡辺2018] (POSデータとはコンビニなどレジのデータのことです)
- ビットコインの取引パターンとバブルの関係を分析することもあります[Bovet et al.2018]



東京大学 数理・情報教育研究センター 久野遼平 2020 CC BY-NC-SA 19

## 関係性の可視化

- 関係性の可視化にはネットワークや行列が使用されます
  - 行列は4-1-5で詳しく説明します
- 下記はネットワーク図として上場企業同士の関係を描写したものです



上場企業同士の関係

東京大学 数理・情報教育研究センター 久野遼平 2020 CC BY-NC-SA 10

## IoTとロボット

- モノのインターネット(IoT)やロボットの利用も盛んです
  - Internet of Things: インターネット経由でセンサーを結合することです

- 例えばAkerではドローンを飛ばして農作物をモニターするアプリを開発しています(<https://www.akersolutions.com/>)



© 2003-2018 Aker Solutions ASA

- ポストダイナミクス社が作るロボットの例は皆さんも見たことがあると思います(<https://www.youtube.com/watch?v=13il13vDH5w>)

東京大学 数理・情報教育研究センター 久野遼平 2020 CC BY-NC-SA 26

## 農業と土木

- 農業や土木でもIoT機器を使用することが多いです

- 例えば熟練した農家の栽培技術のAIでの再現を目標に、IoT機器と機械学習の技術で上手な水やりを学習し、甘いトマトの栽培に成功した研究者もいます(<https://www.jst.go.jp/kisoken/yonetsu/interview/h29/mineno.html>)



出典：2020 科学技術振興機構(JST) 峰野博史、多様な環境に自律順応できる水分ストレス高精度予測基盤技術の確立

- ドローンで撮影した画像から被災地域の推定を行っている研究者もいます[MIT2019] (<https://www.technologyreview.com/s/614246/ai-image-recognition-improves-disaster-response/>)

東京大学 数理・情報教育研究センター 久野遼平 2020 CC BY-NC-SA 28

その他の教材

# コンソーシアムの教材ポータルサイト

## 1. 社会におけるデータ・AI利活用

### 1-1. 社会で起きている変化

キーワード	数式なし または 少なめ	数式あり	補助教材	教科書シリーズとの対応
	<a href="#">社会で起きている変化(スライド・東京大学)</a>			
ビッグデータ、IoT、AI、生成AI、ロボット	<a href="#">データサイエンスの役割(1)(動画・滋賀大学)</a> <a href="#">ビッグデータとIoT/CPS(動画・筑波大学)</a>			<a href="#">教養としてのデータサイエンス</a>
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	<a href="#">データサイエンスの役割(2)(動画・滋賀大学)</a>			<a href="#">教養としてのデータサイエンス</a>

<http://www.mi.u-tokyo.ac.jp/consortium/e-learning.html>

- モデルカリキュラムに対応した教材の収集・公開
  - 東京大学, 東京都市大学, 滋賀大学, 九州大学, 筑波大学, 北海道医療大学
  - 動画, スライド, 演習用事例およびデータ, サンプルコード, etc.
- 教科書「データサイエンス入門」シリーズ (講談社)

まとめ

# まとめ

- 数理・情報教育センター（東大）で開発した教材
  - 講義スライド形式の教材
    - モデルカリキュラムに完全準拠
    - すべてのキーワード（知識・スキル）を解説
    - ページ単位で，教育のために自由に利用可能
  - データ解析例，サンプルコード， etc.
- 数理・データサイエンス・AI 教育強化拠点コンソーシアム
  - 教材ポータル
  - データサイエンス入門シリーズ（教科書）