

2024(令和6)年度

データサイエンス入門 学習ノート

PART 1

内容

PART 1

- 0. #01 準備：データ・情報と2進表現 1
- 1. #02 導入：社会で起きている変化 5
- 2. #03 導入：社会で活用されているデータ 9
- 3. #04 導入：データとAIの活用領域 16
- 4. #05 導入：データとAI利活用のための技術 20
- 5. #06 導入：データ・AI活用現場 31
- 6. #07 導入：データ・AI利活用の最新動向 34

PART 2

- 7. #08 基礎：データを読む 38
- 8. #09 基礎：データを説明する 47
- 9. #10 基礎：データを扱う 52
- 10. #11 心得：データ・AIを扱う上での留意事項 55
- 11. #12 心得：データを守る上での留意事項 61

長崎総合科学大学

データサイエンス入門 #01 0. データ・情報と2進表現

総合情報学部
日當明男



0.1. データと情報の概念 (1/2)

データ・・・ただの[]や[]など。
または、それらをまとめたもの。

データの解釈は、[]次第

データ自体に
[]は
付いていない

例: データ「123」

- 解釈例 { 数「123」(ヒャク・ニジュウ・サン)
- { 数字列「123」(イチ・ニ・サン)
- { :



0.1. データと情報の概念 (2/2)

情報・・・人の[]に影響を
与える[]モノ。自分の[]や
外からの[]

情報の価値 (影響の大きさ) は、
[]次第

情報自体に
[]は
付いていない

例: 情報「明日のテストではコレが出る」

- 受け手側の価値 { 大: 全然準備していなかった人
- { 中: 準備していた人
- { 小: 結構準備していた人



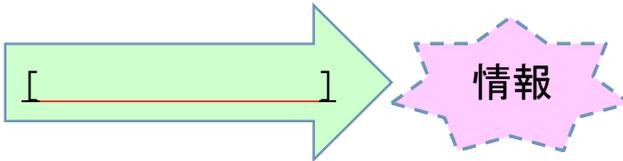
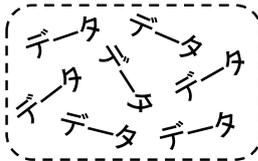
#01-04

0.2. データと情報の処理 (1/3)

データ処理・・・何らかの[]を目的として、
データ(群)に対して[]が行う操作

「同種のデータの集まり」を意識するとき、データ「群」と呼ぶことがある。

データ(群)の[]や[]を求めたり、[]など



NAS 長崎総合科学大学

4

#01-05

0.2. データと情報の処理 (2/3)

情報処理・・・情報に対する3操作

[], [], []
この操作は、[]が行う。

情報処理の主体は、[]。
コンピュータは、[]を人間が作った[]に従って処理する、便利な道具。

ここは、次の項で説明する。

データ処理における[]や[]と同様

NAS 長崎総合科学大学

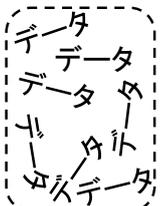
5

#01-06

0.2. データと情報の処理 (3/3)

データサイエンス

・・・何らかの目的をもって、
データに対して適切な操作を施し、
当初の目的に即して、
その結果に対して適切な操作を施す一連の取組



データサイエンス = [] + []

例：データ解析：複数のデータ分析 + 総合解釈



NAS 長崎総合科学大学

6

0.3. データ・情報の表現(1/1)

データ・情報の表現要素

数、文字、記号、絵(画像)映像(動画)、音

このままでも、人間なら
処理できる。

でも、コンピュータに処理を手伝わせたい

◆ [] で扱える表現に変換 ([])

◆ 処理のための []

[] 表現

[] が [] のようになる

0.4. 2進表現の考え方(1/3)

- 2進表現…2つの記号(0,1)の複数桁で、
違い(異なる状態:情報)を表わす方法。
- ビット(bit: binary digit)…2進表現の []。

桁数n (ビット数)	状態数 2^n	2進表現	0以上整数 (マイナスを含む)
1	$2=2^1$	0, 1	0~1 (なし)
2	$4=2^2$	00, 01, 10, 11	0~3 (-2~1)

0.4. 2進表現の考え方(2/3)

桁数 (ビット数)	状態数	2進表現	0以上整数 (マイナスを含む)
3	$8=2^3$	0XX… [] 種類, 1XX… [] 種類	[] (-4~3)
4	$16=2^4$	0XXX… [] 種類, 1XXX… [] 種類	[] (-8~7)
8	$256=2^8$	0XXXXXXXX… 128種類, 1XXXXXXXX… 128種類	[] ([])

8ビットをまとめて、 [] と呼ぶ

0.4. 2進表現の考え方(3/3)

表現要素	2進表現ルール等
数	0以上の整数。マイナスを含む整数。浮動小数点表示の2進表現化 小数点が付く数
文字, 記号	文字コード(ASCII、Unicode、Shift-JIS、EUC等) 記号と2進表現の対応表
絵(画像)	画素ごとの3原色の強さ(0以上の数)の組全体
映像(動画)	少しずつ異なる静止画像の組
音	1秒間に約44,000回の音の強さ(0以上の数)の組

0.5. データ・情報の量(1/2)

- データ・情報の量(サイズ)は、2進表現されたときの桁(ビット)またはバイト数で測る。
- 情報量の単位【b:[]、B:[]】
- 接頭辞

接頭辞	K	M	G	T	P	E	Z	Y
読み方	キロ	メガ	ギガ	テラ	ペタ	エクサ	ゼタ	ヨタ
(約)?倍	千	百万	十億	一兆	千兆	百京	十垓	一杼
(約)10 ⁿ 倍:n	3	6	9	12	15	18	21	24

0.5. データ・情報の量(2/2)

- テキストp.4の例
 10分毎に1枚の写真画像を一生(80年)保存する。
 1枚の写真画像(1700×980画素、フルカラー)
 $1700 \times 980 \times 3B = 4,998,000B \div [\text{ }]$
 10分間隔で80年
 $(60 \div 10) \times 24 \times [\text{ }] \times 80 = 4,207,564.8$
 よって、
 $5MB \times 4,207,564.8 = 21,037,824MB \div [\text{ }]$

データサイエンス入門 #02

1.1. 社会で起きている変化

総合情報学部
日當明男



1.1.0. Society5.0時代に生きる(1/2)

Society1.0: 狩猟社会



Society2.0: 農耕社会



Society3.0: 工業社会



Society4.0: 情報社会



1.1.0. Society5.0時代に生きる(2/2)

Society5.0: []

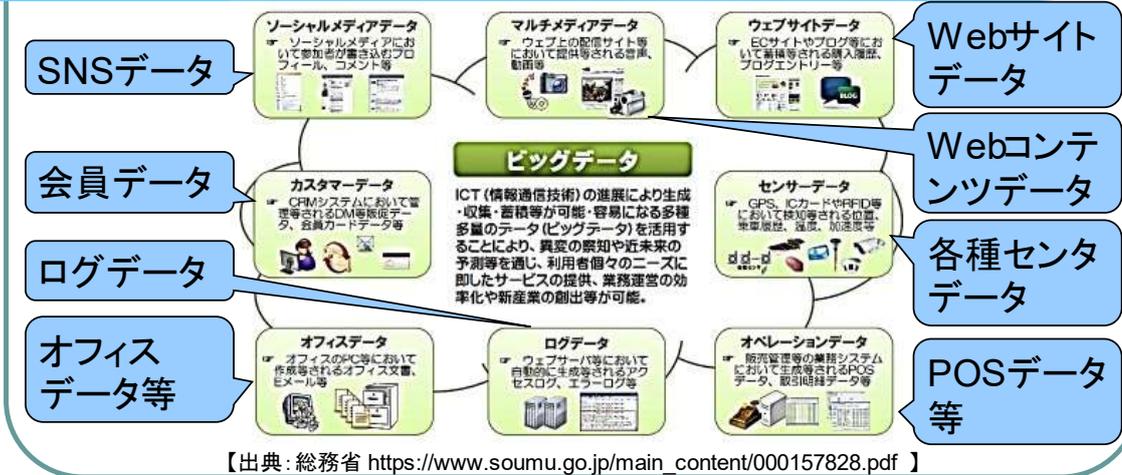
サイバー空間(仮想空間)とフィジカル空間(現実空間)を高度に融合させたシステムにより、経済発展と社会的課題の解決両立する、[]の社会(Society)

【出典: [Society 5.0 - 科学技術政策 - 内閣府 \(cao.go.jp\)](http://cao.go.jp)】

[[Society 5.0ビッグデータ連携がもたらす未来社会像: 動画](#)]



1.1.1. ビッグデータ(1/2)



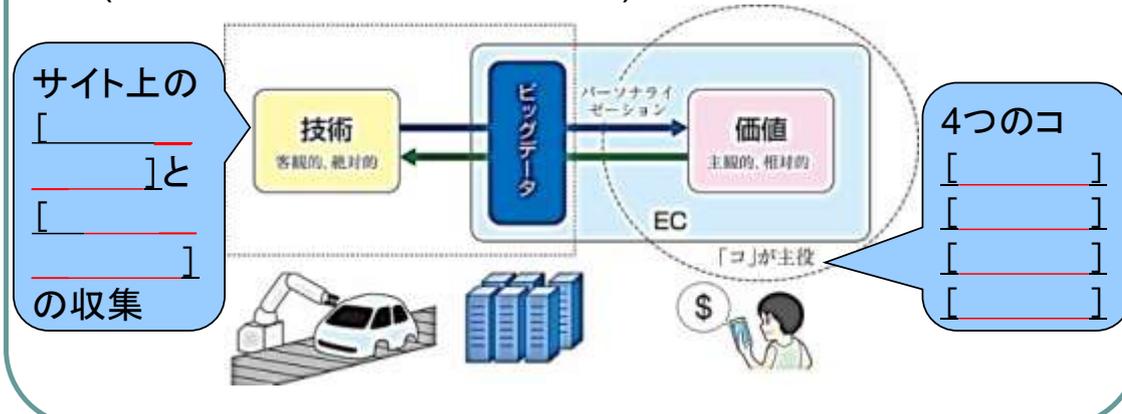
1.1.1. ビッグデータ(2/2)

● ビッグデータの特徴(3V+2V)

- Volume ([]): データの量が多い
- Velocity ([]): データ計測が速い → 記録が速い
- Variety ([]): データの種類がさまざまである
- Veracity ([]): 偽物ではない事が求められる
- Value ([]): 活用するからこそ価値をもたらす

1.1.2. 検索エンジンとSNS(1/2)

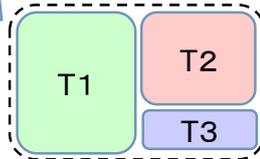
EC(電子商取引: ネットショッピング)におけるビジネス形態



1.1.2. 検索エンジンとSNS (2/2)

● データ駆動 (データドリブン)

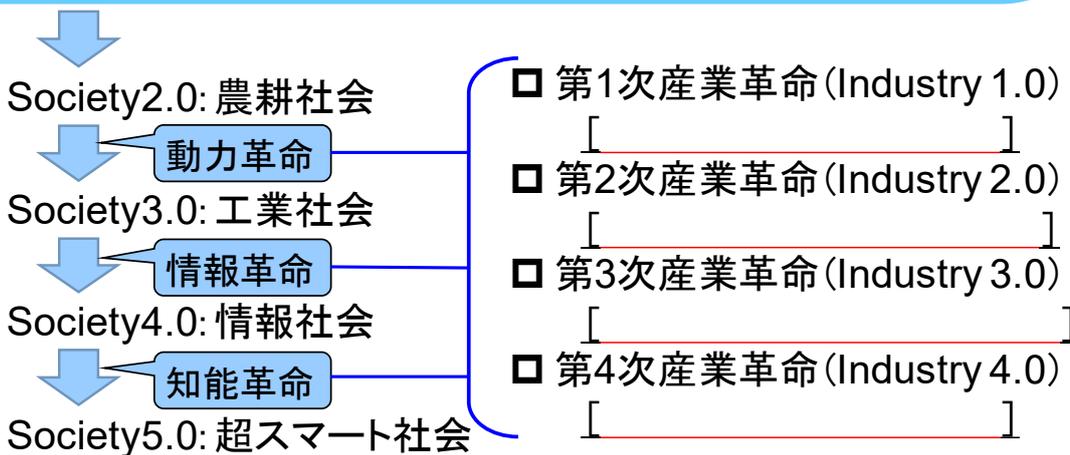
ビッグデータのパターン分類を基礎とする []



収集されたデータ(群)には、
T1 > T2 > T3
の順で [] があった。

今後収集されるデータにも、
似た [] が []
できるので、...

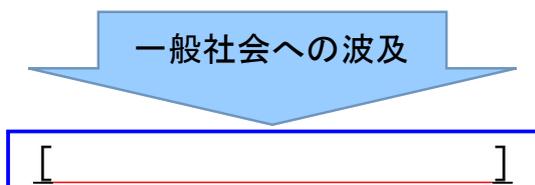
1.1.3. 第4次産業革命 (1/2)



1.1.3. 第4次産業革命 (2/2)

第4次産業革命 (Industry 4.0)

大量の [] から取得される [] の
分析を基に、工場やものづくりの現場の課題解決をする。



データサイエンス入門 #03 1.2. 社会で活用されているデータ

総合情報学部
日當明男

NAS 長崎総合科学大学

1.2.0. AIとビッグデータ(1/1)

- AIとビッグデータ
ビッグデータはAIの学習教材
- ビッグデータの特徴(3V+2V)
 - Volume ([]): データの量が多い
 - Velocity ([]): データ計測が速い → 記録が速い
 - Variety ([]): データの種類がさまざまである
 - Veracity ([]): 偽物ではない事が求められる
 - Value ([]): 活用するからこそ価値をもたらす

NAS 長崎総合科学大学

2

1.2.1. データの種類(1/9)

- a. 調査データ
[] するために収集されたデータ
- b. ログデータ
[] に記録・蓄積されたデータ
- c. 実験データ
実験の過程で [] で収集されたデータ
- d. 観測データ
様々な [] によって収集されたデータ

NAS 長崎総合科学大学

3

1.2.1. データの種類(2/9)[a.調査データ(1/3)]

事例	取得方法	取得組織
国勢調査、政府統計データ	[]	政府
民間の信用調査	[]	民間企業

● 調査目的

国勢調査、政府統計データ

…国内の[]や[]を知るため。

民間の信用調査

…投資や貸し付けの[]を判断するため。

1.2.1. データの種類(3/9)[a.調査データ(2/3)]

● アンケート調査が持つ課題

1. 如何に、データを得るか。

[]…近年では、不在や拒否が増えている。

[]…回収に時間がかかる。回収率が低い傾向にある。

[]…回答に必要なWeb操作能力に依存する。

2. 誰からデータを得るか。

[]が問題

[]…全体に対する調査

([] に課題)

[]…抽出した一部に対する調査

([], 結果の [] に課題)

1.2.1. データの種類(4/9)[a.調査データ(3/3)]

● アンケート調査データにおける留意点

● コスト

例	データ回収・管理	データ利用
国勢調査、 政府統計データ	[]	低い
民間の信用調査	[]	高い

● 結果の信用性

[]に[]があれば、結果の信用性は低い。

1.2.1. データの種類(5/9)[b.ログデータ(1/3)]

事例	取得方法	取得組織
[]ログ	検索エンジン	Yahoo、Google等
[]ログ	ECサイト	アマゾン、楽天等
[]ログ	スマホ	ドコモ等の通信会社
プローブデータ	カーナビ	自動車メーカー
機器の稼働ログ	機器内のIoTセンサ	機器メーカー

- ログデータの特徴
自動収集されるので、[]化しやすい。

1.2.1. データの種類(6/9)[b.ログデータ(2/3)]

- 検索ログの利用
検索キーワードから、利用者の[]の傾向を知る。
- EC購買ログの利用
利用者の購買履歴や検索履歴から、[]の傾向を知る。
また、似た利用者データから、[]を提示する。
- 人の行動ログの利用
スマホのハンドオーバー(基地局自動切換え)機能による、基地局内の接続スマホの[]から、
スマホ保持者の[]を推測する。

1.2.1. データの種類(7/9)[b.ログデータ(3/3)]

- プローブデータ
走行車両に搭載された[]のデータ
⇒ 東日本大震災直後の[]の確認に活用
- 機械の稼働ログ
各種機器に搭載されたセンサーの[]データ
⇒ 故障前の兆候を捉えて、[]を予告

1.2.1. データの種類(8/9)[c.実験データ(1/1)]

実験時の計測装置等が取得するデータ

- 昔は、[]が記録し、分析していた。
- 今は、[]に記録され、データのサイズも大きくなり、実験者が分析・判断する[]。
⇒データの[]



データの分析と判断のための[]を自動的に提示する[]が必要

1.2.1. データの種類(9/9)[d.観測データ(1/1)]

各種観測機器等が取得するデータ

例えば、

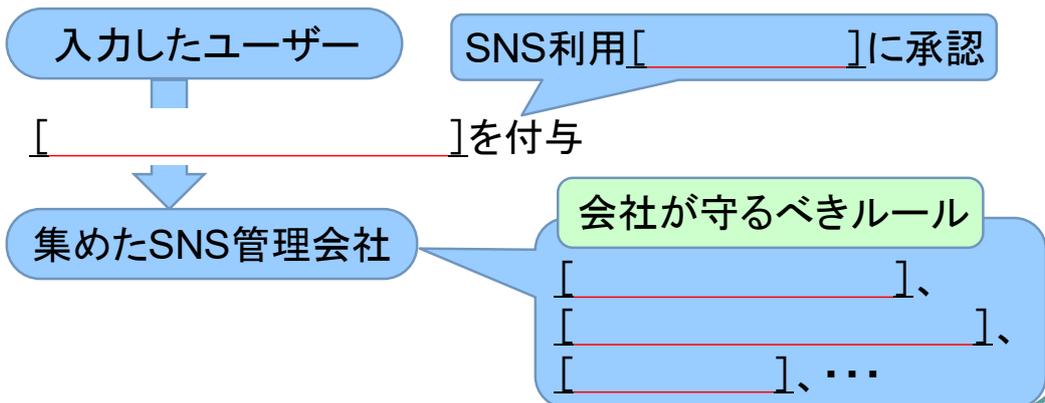
- []の観測データ
- スーパーカミオカンデの[]観測データ
- []の天体の明るさ(電波強度)観測データ



データの分析と判断のための[]を自動的に提示する[]が必要

1.2.2. データの所有者(1/3)

- SNS上のデータは誰のもの？



1.2.2. データの所有者 (2/3)

- データの種類
 - 1次データ: []で集めたデータ
 - 2次データ: 公共機関や[]が集めたデータ
公共機関が集めたデータの一部が公開されて
[]と呼ばれる。
 - 3次データ: 複数のデータを[]
使いやすくしたデータ
 - メタデータ: データ取得時や取得データの
[]データ

1.2.2. データの所有者 (3/3)

- データの信頼性
 - 計測誤差: 計測機器に[]する誤差
- 実験・観測データには必ず誤差がある。
- 観測誤差: 観測時の[]の誤差
 - []をある程度
[]できる
- []誤差: 測定機器や測定方法の癖によるものなど。
 - []誤差: 測定時にランダムに含まれる誤差。
 - []誤差: 測定者の経験不足や不注意による誤差。

1.2.3. 構造化データと非構造化データ (1/6)

- 1件の[]データの姿の例
 - 気温データ・・・[]
 - 地球上の位置データ・・・緯度と経度の[]
 - 学籍番号データ・・・コースID、入学年度、コース内番号の
[]
 - 人データ・・・名前、生年月日、身長、体重、、、等の
文字記号、日付や値などの[]

1.2.3. 構造化データと非構造化データ(2/6)

- []データの仕組み
[]付けられた複数の項目の組み(構造)を持つデータ



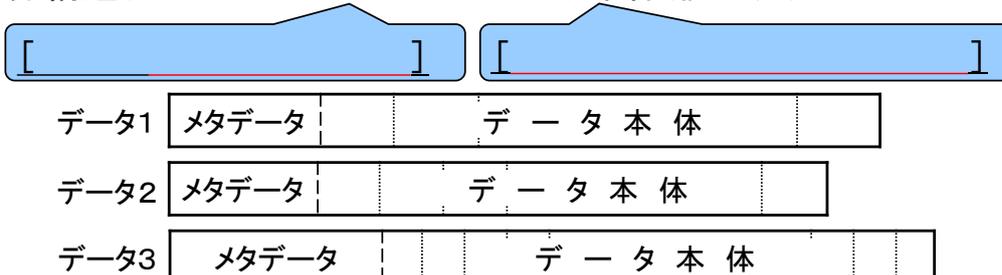
1.2.3. 構造化データと非構造化データ(3/6)

- 1件の []データの姿の例
 - テキスト(文章)データ...個々の文字(2進)の列。
説明情報: []
 - 音声データ...サンプリング周期毎の音の強度値の列。
説明情報: []
 - 静止画像データ...各画素の色情報の画素数個の列。
説明情報: []
 - 動画データ...一定周期毎の静止画像情報の列。
説明情報: []

1.2.3. 構造化データと非構造化データ(4/6)

- []データの仕組み
[]とセットにしたデータ。

非構造化データ = メタデータ + データ本体(値の列)



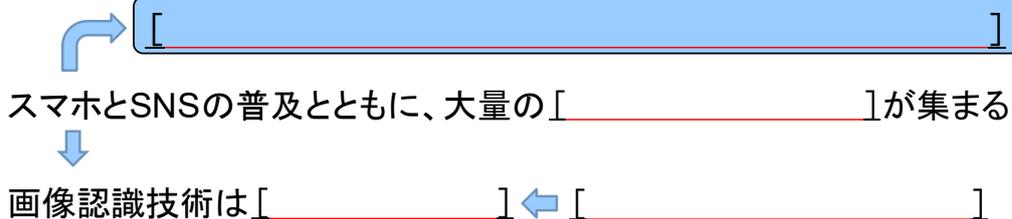
1.2.3. 構造化データと非構造化データ(5/6)

- 画像認識
 画像中の「物体は何か」を[] 答える作業
 - ・特定物体認識・・・[] 物体の認識
 - ・一般物体認識・・・[] 物体の認識
- ◎ セグメンテーション
 複数の物体が入っている画像の中から、
 [] を切り出す作業。

同じ領域に複数の物体があると [] ので

1.2.3. 構造化データと非構造化データ(6/6)

- 深層学習
 多くの中間層を持つニューラルネットワークによる学習方式
- 深層学習による画像認識
 (領域内の)すべての画像に[] をつける作業が必要。



1.2.4. 自動翻訳(1/1)

- 自然言語の機械での翻訳
 - ▼ 17世紀: 同意語への記号の割り付け
 - ▼ 1954年: 特定の構文の翻訳…ロシア語は、[]
 - ▼ 1964年: [] レポート…機械翻訳の [] を指摘
 - ▼ 1980年代: [] の機械翻訳… [] は無理
 [] 等の翻訳ルールを [] 作成
 - ▼ 1990年代: [] 機械翻訳…精度は [] に向上
 [] な翻訳機モデルを [] が構築
 - ▼ 2010年代: [] 機械翻訳…精度は [] に向上
 翻訳機モデルを [] で構築

データサイエンス入門 #04 1.3. データとAIの活用領域

総合情報学部
日當明男



1.3.1. 事業活動におけるデータ・AI活用(1/7)

バリューチェーン: 企業活動における価値の連鎖



1.3.1. 事業活動におけるデータ・AI活用(2/7)

● 企画・研究活動における利活用

□ 企画・計画

的確な製造・販売計画構築のために、データに基づく[]。

□ 新技術の開発、新素材・新薬の推測

過去の実験データや人間の適切な判断を基にした []や[]を活用。

□ 新商品推測

アンケート等の調査を基に、[]を活用。



1.3.1. 事業活動におけるデータ・AI活用(3/7)

- 購買物流・調達における利活用
 - []に基づいて、適切な []と[]を立て、トータルコストの削減を図る。
 - トータルコストの削減 []、 []について、過去のデータを用いる
 - 原材料価格の予測 []と[]から推測



1.3.1. 事業活動におけるデータ・AI活用(4/7)

- 製造におけるデータ・AI利活用
 - 品質管理や製造工程の改善 []や工程の []の発見に過去のデータを活用する
 - 製造機械の保守 []に、過去のデータやAIを活用する
 - 生産計画や検査工程への支援 熟練工の []の一部の自動化



1.3.1. 事業活動におけるデータ・AI活用(5/7)

- 出荷物流における利活用
 - 倉庫内作業の効率化 []を考慮した []の最適化
 - 輸送業務の効率化 []に基づいた []の適切な配置



1.3.1. 事業活動におけるデータ・AI活用(6/7)

- マーケティング・販売における利活用
需要予測のための[]
と広告の効果を調査する。
- 市場動向調査、POSデータの提供
[]だけでなく、幅広くデータを収集。
ポイントカードで個々人の[]を把握。
- 広告効果調査
既出広告の認知度や購買への効果を踏まえて、新規広告の[]や[]を提案。



1.3.1. 事業活動におけるデータ・AI活用(7/7)

- サービスにおける利活用
- アフターサービス
[]にチャットボットを活用。
- 保守サービス
[]や[]を
基にしたサービス



1.3.2. 活用目的ごとのデータ・AI活用(1/4)

- 仮説検証
P(仮説設定)→D(調査)→C(仮説検証)→A(必要な改善)にて、
Dにおける十分なデータを基に、Cを行う。
調査内容は、[]に役立つように決める。
- 知識発見
既に取得したデータから[]を発掘する。
例：レシートデータから、[]の
[]の傾向が見つかった。

1.3.2. 活用目的ごとのデータ・AI活用(2/4)

- 原因究明
故障・不良品・顧客離反への[]をデータから発見。
このためには、[]のデータが必要。
製造段階の特定には、[]を高める。
- 計画策定 []とも言う
シフト管理: 最適な勤務[]を見出す。
配送計画: 配送経路の[]、トラックの[]
や運転手の[]の適切化
人員配置: 業務量や危機対策に応じた[]

1.3.2. 活用目的ごとのデータ・AI活用(3/4)

- 判断支援
営業判断: 営業利益の[]
医療診断: 画像による[]の提示
防災: 各種データを用いた[]
スポーツ: []の分析
- 活動代替
自動運転: 人間の的確な[]を学習
無人店舗: 顧客の[]の把握に活用
自動収穫・選別: 作物の[]に活用

1.3.2. 活用目的ごとのデータ・AI活用(4/4)

- 新規作成
作品制作: 作者の作風を学習させたAIが
[]を作る。
自動生成: 新聞記事・ニュース記事や
天気予報から[]を自動生成する。



【参考】

[自動要約 無料ツール by ユーザーローカル \(userlocal.jp\)](https://userlocal.jp)

データサイエンス入門 #05

1.4. データとAI利活用のための技術

総合情報学部
日當明男

NAS 長崎総合科学大学

1.4.0. 利活用における留意事項(1/1)

- データ解析
[]データから、[]を引き出す事
[]、[]、[]
- データの可視化
データ(群)の[]ための表現方法
- AIへの正しい理解
AIに、[]は無い。
AIは、[]によって賢くなる。
AIは、[]に対しては期待できない。

NAS 長崎総合科学大学

2

1.4.1. 誰もが無意識に・・・(1/1)

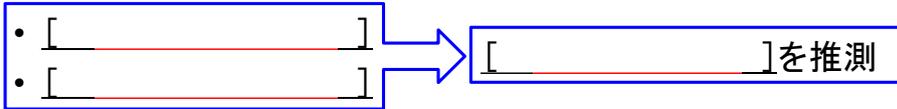
- 予測
 - スパイシーな香りがするので、今晚は[]だろう。
 - この人なら、この商品も[]だろう。
- 傾向の発見
 - 昼食後には、[]ことがよくある。
 - A商品を買った人の多くは、[]している。
- 分類(グループ핑グ・クラスタリング)
 - あ的那个人は、[]。
 - 関西には[]派、関東には[]派の人が多し。

NAS 長崎総合科学大学

3

1.4.2. さまざまなデータ解析－予測(1/3)

● 予測



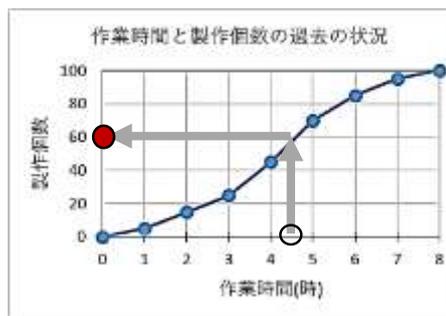
手法	内容	難しさ
手法1	過去の変化が [] と想定	過去の変化と同じ変化が []
手法2	過去の変化が [] を参考	過去の変化が類似しても、その後の変化が []

1.4.2. さまざまなデータ解析－予測(2/3)

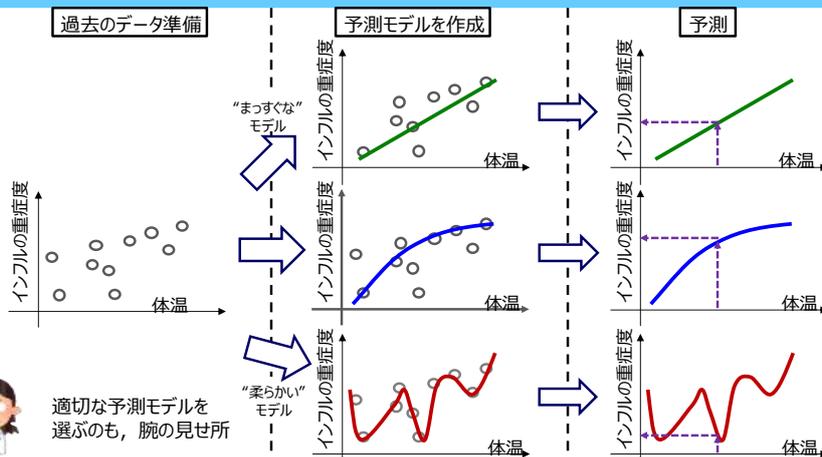
● 補間

実際には [] のデータを、観測した [] のデータから推測する

例(右図)
4.5時間の作業時間の場合、
だいたい [] 個の製品を
製作できそうだ。



1.4.2. さまざまなデータ解析－予測(3/3)

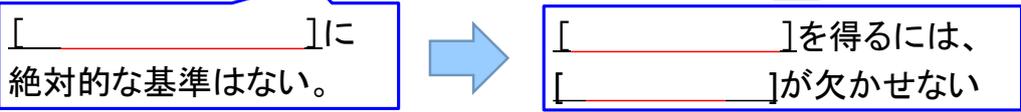


適切な予測モデルを選ぶのも、腕の見せ所

1.4.3. さまざまなデータ解析－分類(1/1)

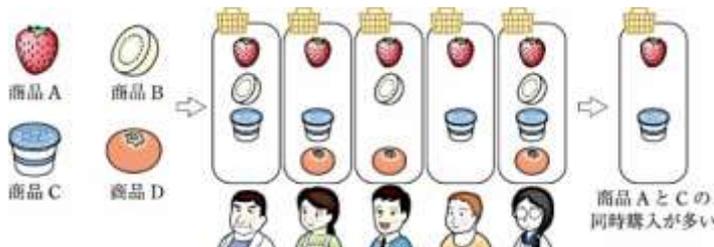
データ内の[]や[]に注目

- グルーピング: 出来るグループが事前に[]
まとめる基準:[]
- クラスタリング: 出来るグループが事前に[]
まとめる基準:[]



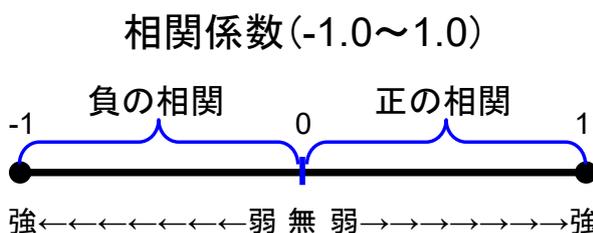
1.4.4. さまざまなデータ解析－発見(1/5)

- マイニング(発掘)
大量のデータに埋もれている[]を見つけ出す
- 頻出パターンの発見
高い確率で[]する項目値の組合せを見つける

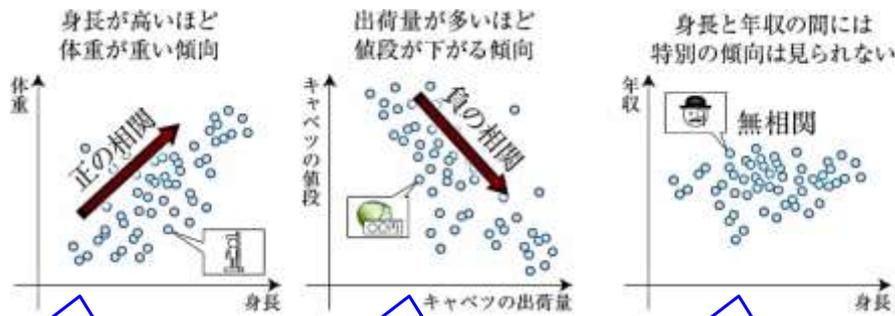


1.4.4. さまざまなデータ解析－発見(2/5)

- 相関分析 構造化データに対する分析
データ内の2つの項目間の[]を見出す手法
相関の強弱: 増減関係の[]の強さ



1.4.4. さまざまなデータ解析－発見 (3/5)



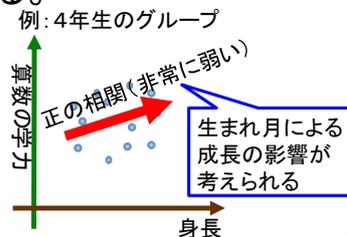
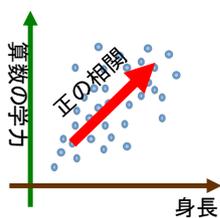
増減関係の傾向が []

増減関係の傾向が []

増減関係の傾向が []

1.4.4. さまざまなデータ解析－発見 (4/5)

- 疑似相関: 隠れた [] が重要 [] 関係がある2項目も、 [] を介して [] に関連しているだけかも。 [] で分類したグループ内では、関係が [] 事もある。



1.4.4. さまざまなデータ解析－発見 (5/5)

相関と因果は全く異なる関係

- 相関関係
2項目間の [] の関係
- 因果関係: 相関関係の []
2項目間の [] の関係



[], []
等によって調べる。

項目Bの値が [] となり、
その結果として、
項目Yの値が []。

1.4.5. データ解析の関連話題(1/1)

- 最適化

[]内で、[]の
[]を図る。

例:限られた予算内で、皆が喜ぶ
景品をたくさん購入する。

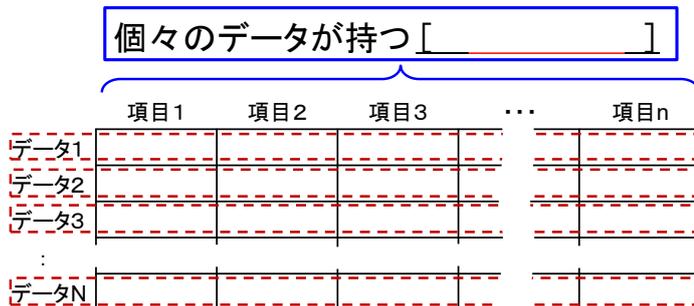
- シミュレーション **「シュミレーション」ではない!**

[]の[]を模擬した
[]を用いて行う実験。
その精度は、[]によって
[]していく。

1.4.6. 非構造化データ処理(1/9)

- []データ

順序付けられた複数の項目の組み(構造)を持つデータ



1.4.6. 非構造化データ処理(2/9)

- 構造化データの例

人の体格データ

	体重	身長	肩幅	腕の長さ	足の長さ
Aさん	45.8	163.5	38.6	42.0	69.5
Bさん	68.4	170.6	40.5	45.2	72.2
Cさん	76.2	168.4	41.6	43.9	70.8

- []データ

構造を持たないデータ。

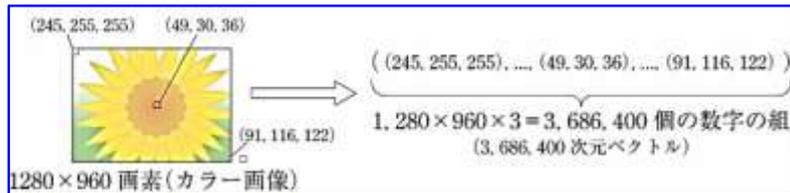
例:画像(静止画像、動画像)、[]、[]等

1.4.6. 非構造化データ処理 (3/9)

画像データ

- 画像データのデジタル化

画像全体を格子状に分割した[]ごとの色情報(光の3原色:[], またはグレイ度)を全画素に亘(わた)って組み合わせて表す



フルカラーの場合

3バイト/画素
=3.686MBの
画像データ

1.4.6. 非構造化データ処理 (4/9)

- 動画データのデジタル化

1秒間あたり複数枚の静止画像([])を []に記録したデータ

例: 1秒間に30フレーム(4MB/フレーム)で1.5時間の動画

$$4\text{MB} \times [] \times [] \times [] = 648,000\text{MB} = 648\text{GB}$$

1秒

1分

1.5時間

BD(25GB)26枚分

とても、持ち歩き出来ない

1.4.6. 非構造化データ処理 (5/9)

- データの圧縮

バイトやビットを小さくすること[]。
データの表現を[]して、サイズを小さくする。

圧縮法

- []圧縮: 完全に復元される
LZH、ZIP、PNG、GIF、Huffyuv、mp3HD、...
- []圧縮: 多少劣化して復元される
JPEG、MPEG(-1,2,4)、MP3、...

比較的、圧縮率が高い

1.4.6. 非構造化データ処理(6/9)

- コンピュータビジョン
[]をコンピュータで実現する分野
 - 画像内の複数物体の[]
 - 複数枚の2D画像から3Dモデルを[]
 - 画像中で動くものを[]
- 画像認識
[]によって、急激に発達

自動運転や
画像監視など
に活用

「夏休みの宿題(1966年)」の解答(参考:1.4.8(2/2))

1.4.6. 非構造化データ処理(7/9)

- 音データのデジタル化
[]の強さを1秒当たり[]、
[]に記録したデータ。
- ※標本化定理
サンプリング周波数の半分の周波数までの信号は復元可
- 音データの処理
音声処理:[]、[]など
音楽処理:[]、[]、[]など
環境音処理:[]、[]

サンプリング周波数: 約44,000Hz

1.4.6. 非構造化データ処理(8/9)

- サンプリング周波数^(*)が44000ヘルツのPCMデータとして音声を記録した時、1分間の音声データは何バイトになるか。
 - (*) 周波数[ヘルツ]=1/周期[s]: 1秒間内の周期の回数
 - 1分は[]秒。 単位: 1秒あたり? 回
 - サンプリング周波数44000ヘルツでの計測は、[] 回/秒。
 - PCMデータでは、1回の計測で[]バイト(65536段階)使用。
- []バイト × []回/秒 × []秒/分
= 5,280,000バイト/分 ≒ 5MBバイト/分

1.4.6. 非構造化データ処理 (9/9)

- 文章データのデジタル化
文字要素(記号)にコード([])を割り振り、出現順に並べる(符号化)。
[]: Shift-JIS、Unicode、EUC、...
- []
[]と[]の際の文字コードの違いで生じる
- 自然言語処理(人間が用いる言語に対する処理)。
[]、[]、[]、センチメント分析(感情分析)など

1.4.7. データ可視化 (1/3)

データ(群)の[]ための表現方法。

● データの可視化手法例

いくつかの特徴は、
[]なる。

- グラフ
データ内の[]や項目値が同値(同級)の[]を比較できる
- 地図を用いた可視化
国や地域ごとの[]の状況を地図上に表示

1.4.7. データ可視化 (2/3)

● グラフ

● 棒グラフ・円グラフ

棒グラフ: データ間での[]やデータ群内での[]を見る

円グラフ: 項目値やデータ数の[]に対する[]を見る

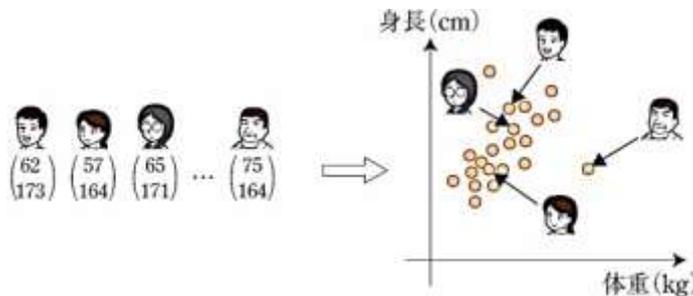
● 折れ線グラフ

[]に並ぶ項目の[]を見る。
未観測時点での[]にも適する。

1.4.7. データ可視化 (3/3)

● 散布図

2項目の値の組を[]として、平面に配置する。
2つの項目の値の[]を見る。



1.4.8. パターン認識技術 (1/2)

● パターン認識

あるデータを、[]された
[]に振り分ける事。



1.4.8. パターン認識技術 (2/2)

● コンピュータによるパターン認識

1966年 夏休みの宿題(画像内の物体の認識): []
[参考: <http://soqdoq.com/teq画像認識aiの発展20112020/>]

↓
2010年代 DNNにより飛躍的に進展: []

● 最近傍法

[]モノと認識

[]を使用



1.4.9. 人工知能 (1/6)

- 人工知能 (Artificial Intelligence)

[]
 現在では、[]によるAIが主流

- チューリングテスト(1950年)

壁越しの対話で、相手が[]を判断する

批判:
 思考実験『中国語の部屋』

現実には難しい。が、
 最近の[]は
 限られた分野では[]

ChatGPTは[]でパス?

1.4.9. 人工知能 (2/6)

- 中国語の部屋【思考実験】



ルールに基づく[]は、
 「知能」と言えるか!

これまでのAIには
 このタイプが多い。

1.4.9. 人工知能 (3/6)

- 特化型AI: []人工知能
 []にだけ対応した人工知能

近年、登場している人工知能はすべて[]
 人間と同程度の性能に[]ケースも多い

→ 現在のAIは[]。

- 汎用AI: []人工知能
 []に対応した人工知能(未だに定義が不明確)
 研究段階で、当分は[]と思われる(私見)

1.4.9. 人工知能 (4/6)

- AIの能力向上【参考1.1.4. AIの驚異的発達(1/2)】
全てのノードでの[]を、
[]な[]で調整する。

AIの能力向上には、学習データの量と質が []
量: []を活用
質: 人間の []

AIは、与えられた学習データの []しか持てない

1.4.9. 人工知能 (5/6)

- 今のAIにできる事、できない事
【参考: https://aismiley.co.jp/ai_news/whatai-can-not-do/】
 - 出来る事
大量のデータ処理、[]、
[]、「ゲーム対戦」も含む、
[]、画像処理や画像認識など
 - できない事
[]の説明、[]の処理、
[]対応、[]な作業など
- ➡ AIも[]。 []がある

1.4.9. 人工知能 (6/6)

- 東ロボプロジェクト(国立情報学研究所:2011年始動)
目標:2016年までに、センター試験で高得点
2021年に東京大学入試突破

↓ 【参考: <https://asagaku.com/chugaku/topnews/8294.html>】

[] 年に続行断念の理由

- ① []、と
[]の明確化
- ② []の []を揃えられない

データサイエンス入門 #06 1.5. データ・AI活用現場

総合情報学部
日當明男



1.5.1. データ分析による意思決定 (1/5)

意思決定の変化([] ⇒ [])
[] の排除がポイント

- SABRmetrics(セイバーメトリックス)
[] に基づいた野球選手の評価指標
映画: マネーボール (<https://eiga.com/movie/55274/>)



1.5.1. データ分析による意思決定 (2/5)

データ分析の3局面(図1.5.1)

局面	目的	手法の例
説明的データ分析 	何が起きたか知りたい 	可視化, データマイニング, クラスタリングなど
予測的データ分析 	何が起きるか知りたい 	回帰分析, サポートベクター マシン, 深層学習, 時系列 解析など
指示的データ分析 	何をすればよいか知りたい 	線形計画法, 凸最適化, 実験計画法, A/Bテスト, ベイズ最適化, 強化学習など

これから学ぶ
かもしれない
分野。



#06-04

1.5.1. データ分析による意思決定 (3/5)

① 説明的データ分析

- 分析前: [] ([])
[]、[] や
[] を [] で調査
- 分析後: []
現材の状況を説明できる [] を構築

NAS 長崎総合科学大学

4

#06-05

1.5.1. データ分析による意思決定 (4/5)

② 予測的データ分析

- 分析前: []
[] 仮説を設ける
- 分析時
[] を参考に、
今後 [] を予測する
- 分析後: []
[] に沿った [] の構築

NAS 長崎総合科学大学

5

#06-06

1.5.1. データ分析による意思決定 (5/5)

③ 指示的データ分析

- 分析前: []
複数の予測モデルに対して、[] での
[] を構築
- 分析時
複数のシナリオに対する [] を示す。
- 分析後: []
[] を示す



NAS 長崎総合科学大学

6

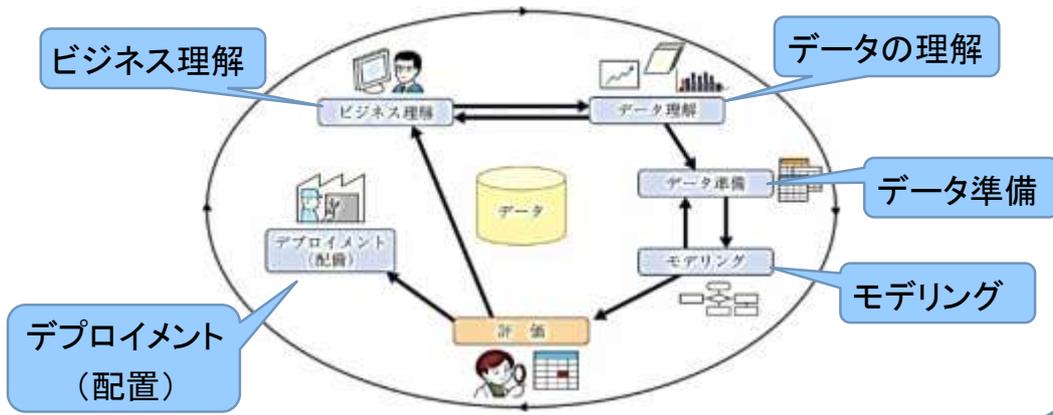
1.5.2. 情報技術による自動化(1/1)

(特化型)人工知能を用いて、ビジネスの効率化を図る

- 第1段階:[]
 情報技術を活用して[]を[]する
 []ビジネスを[]
 []や[]などを活用
- 第2段階:[]
 情報技術を活用して[]を[]する
 []ビジネスを[]
 []や[]などを活用

1.5.3. データ分析・自動化の実際(1/3)

CRISP・DM(業界横断型データマイニング標準プロセス)



1.5.3. データ分析・自動化の実際(2/3)

- ビジネス理解:[]
 データ分析の[]。
- データ理解
 過去に取得されたデータ([])の
 []の確認
- データ準備
 []を想定して、
 []を準備する。

1.5.3. データ分析・自動化の実際 (3/3)

- モデリング
様々な手法での[]
- 評価
[]による評価
良くなかった場合⇒「[]」に戻る
良かった場合⇒「[]」へ
- デプロイメント(配置)
[]する

1.5.4. 組織的考慮点 (1/1)

データ活用による経営課題解決における問題点。

- データ分析やAI活用に詳しい人財を社内に期待できない。

- ↓
- DX技術は[]の源泉なので、外部委託には[]の問題がある。
 - []するか、[]を取り込むか。
 - 育成も取り込みもできない[]は？

データサイエンス入門 #07

1.6. データ・AI利活用の最新動向

総合情報学部
日當明男

1.6.1. AI等を活用した・・・(1/3)

(a) シェアリングエコノミー(共有経済)への活用

- []の予測
- 資源の[]
- 利用者の[]
- 利用状況に応じた[]

※シェアリングエコノミー
インターネットを介して、[]の
[]などのサービスによる経済

1.6.1. AI等を活用した・・・(2/3)

(b) 商品の推薦

ユーザの[]の履歴を基に、
利用者の[]商品を推薦する。

(c) 監視(サーベイランス)

- []
- 店舗内顧客の[]
- 店舗の[]

1.6.1. AI等を活用した・・・(3/3)

(d)DX: 今後の社会の基盤

AIやDXの導入推進 ⇒ 現在の仕事の[]

- 無くなる仕事
[]できる仕事
 - 残る仕事
AIを[]仕事、AIに[]仕事
- ↪ []

#07-05

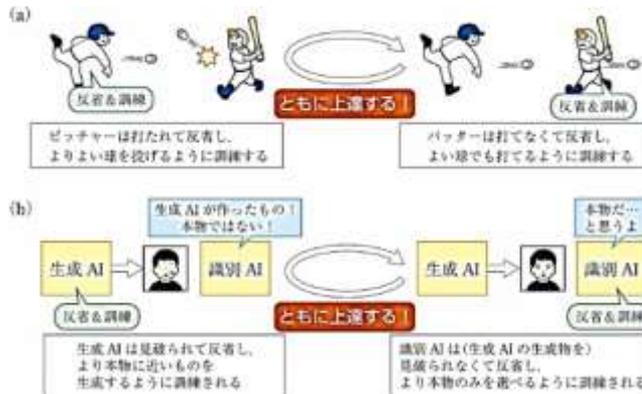
1.6.2. AI最新技術の活用例(1/6)

(a)敵対的生成ネットワーク(GAN)

2つのAI

- []
- []

の[]
的な学習。



NAS 長崎総合科学大学

5

1.6.2. AI最新技術の活用例(2/6)

(b)敵対的生成ネットワークに関する詳細

- []での学習によるモード崩壊
[]学習データが必要
- 様々な社会問題
 - Deepfake: 精巧な[]の悪用
 - TEZUKA2020: 故人の[]
 - 著作権: []の著作権

NAS 長崎総合科学大学

6

1.6.2. AI最新技術の活用例(3/6)

(c)強化学習

- []による学習。
- 人間も行う[]からの学び。

AIがデータから学習して、識別や予測の仕組みを自動的に構築する技術の総称

強化学習

教師あり学習
教師なし学習

教師あり学習
教師なし学習

強化学習

人間の脳を模した学習手法。データの特徴は自ら見出す。

深層強化学習

NAS 長崎総合科学大学

7

1.6.2. AI最新技術の活用例(4/6)

●人間が行っている強化学習



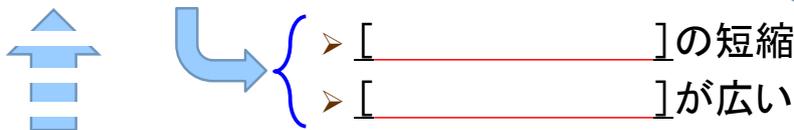
(d)強化学習の応用先

ゲーム、ロボット制御、自動運転、化学、推薦

1.6.2. AI最新技術の活用例(5/6)

(e)転移学習

ある領域での [] に、
[] で学習させる事。



人間の場合

- 一つの [] を取得した人は、他の [] も取得しやすい。
- [] を学んだ学生に、会社で [] 人が多い。

1.6.2. AI最新技術の活用例(6/6)

(f)倫理と社会

AIに [] 判断をいかにさせるか？

【Moral Machine(Awad et al. 2018)】

「ある人を [] ために、
他の人を [] ことは許されるか」

の問題の生成・判断に4000万人が参加

<http://moralmachine.mit.edu/hl/ja>

<https://wired.jp/2019/01/02/moral-machine/?msclkid=ea1b2978cf6511eca50ea3011956e9ec>