

# コンピュータ演習

AIリテラシー 05講 データ・AIの技術

# 目次

- 第05講 データ・AIの技術
  - データ解析とは何をしているのか
    -
  - 可視化の手法にはどう言ったものがあるのか
  - 非構造化データの処理とは
    - 音声合成
    - 歌声合成
  - AIの技術とは

# 第05講 データ・AIの技術

# データ解析とは何をしているのか

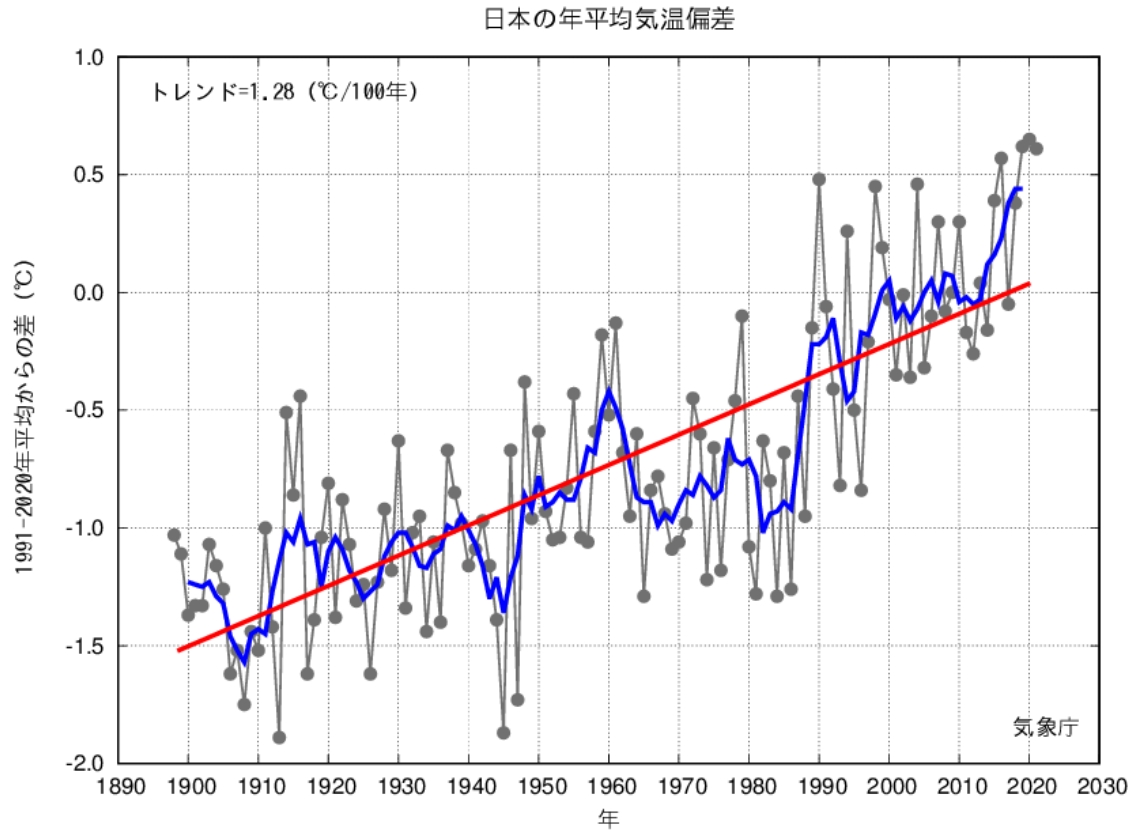
## 昔から解析は行われていた

解析とは「データの有意な規則性を発見する活動である」

昔から行われてきた手法で、

- 情報システム
- AI

を利用することで更なる解析が必要となってきた。



細線（黒）：各年の平均気温の基準値からの偏差

太線（青）：偏差の5年移動平均値

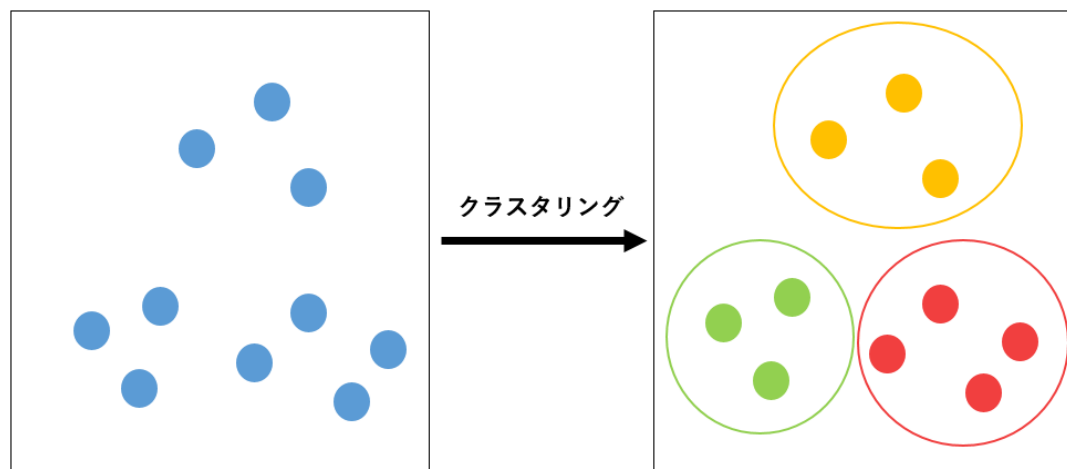
直線（赤）：長期変化傾向。

基準値は1991～2020年の30年平均値。

## クラスタリング解析

クラスタリングとはデータをグループ化すること

難しくいうと、「分類対象の集合を、内的結合 (internal cohesion) と外的分離 (external isolation) が達成されるような部分集合に分割すること」



<https://amplify.fixstars.com/ja/techresources/application/clustering/>

## クラスタリングの難しさ

- どのくらい密集していればクラスターと考えるのか
- バラついててもいいのか？
- 全体として何このクラスターに分けるのか

指針によってかなり異なる分析結果が出てしまう。

データサイエンティスト(データ分析の専門家)の腕の見せ所



## パターン認識(認識)

次のグラフを見てパターンを掴むことはできるでしょうか？

第1波～第7波 感染者数グラフ NHK

パターン認識とは「画像や音声など膨大なデータから一定の特徴や規則性のパターンを識別して取り出す処理のこと」

- 人間でもパターンに気づくことはできる
- 先入観によって人の気づきが邪魔されることがある

## ビールと紙おむつとの売上の例

ビールと紙オムツを並べて陳列したら、両方一緒に買う人が続出して売り上げが増加した

- 先入観として「ビールと紙オムツは普通一緒に買わないのでは？」と思い込んでしまう
- 仮説に沿ってデータを収集解析するため**ビールと消しゴム**でどうかについては気がつかない

## AIによるパターン認識

- 複雑なパターンを発見できる
- 人間では盲点になって気付きにくいパターンを発見できる

ビッグデータとAIを組み合わせれば

- 人間の発想では思いもよらなかった目新しい発見ができるかも
- 仮説を立てなくても相関を見つけることができる

## AIも万能ではない

- 収集したデータに偏りがあれば間違った判断を下す
- 何の役にも立たない発見かもしれない

チェスや将棋でもAIとデータの扱いに長けた人間を組み合わせることで、AI単体よりも強くなることが知られている。

**可視化の手法にはどう言ったものがあるのか**

## グラフによる可視化

- データをわかりやすく見せる**可視化**の重要性が増している
- 一方、可視化で誇張することで誤ったメッセージが伝わってしまうチャートジャンク・詐欺グラフが問題になっている

チャートジャンク：グラフを用いて情報を伝える際、グラフ中の過剰なデザイン、3Dや、シャドー効果、太線のグリッド等のビジュアル要素を適切に用いない事により、本来伝えたい情報が正しく伝わりにくくなったり、情報が歪められて伝わったりする現象のことである。

## パレート図

あるものを構成する項目ごとの値、あるいは階級ごとの度数を大きい順に並べたものと、その累積の構成比を表す折れ線グラフを組み合わせたグラフ

全体の中で大きな影響を占めるものが何であることを明確にし、重要な問題を特定するための手法

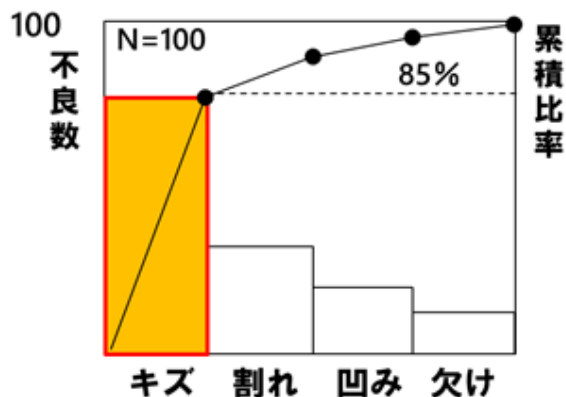
## 第2章：パレート図とは

### 1. パレート図と重点指向

全体の中で大きな影響を占めるものが何であることを明確にし、重要な問題を特定するための手法。

#### パレート図の主な使用目的

- 重点的に取り組む問題を特定する
- その影響がどの程度か把握する
- 改善前と改善後の効果を確認する



不良を要因別に分類した例

**全体の中でキズが85%！  
キズを重点的に対策すべき！**



- パレート図は、各分類項目の影響度が一目で分かるようになる！
- 重点指向に欠かせない手法！！



## 地図を使った可視化

場所によるデータを地図にプロットするタイプのもの

- [JapanTaxi](#)
- [Kepler.gl](#)

## 動的な可視化

見ているうちに情報が変化する動的なデータを可視化するもの

- [FireEye](#)
- [CyberThreat Realtime Map](#)

## 非構造化データの処理とは

言語処理・画像処理・音声処理といった非構造化データの処理も高度化した

## 言語処理

- かなり自然な翻訳が行えるようになってきている
- 文章の意味や構造を把握しているわけではない
- ビッグデータを活用し、「この単語の次にはこのイディオムが来ることが多い」などを併用することによって実用的な文章を組み立てられるようになっている。

## 画像処理

画像をPhotoshopなどで長いことレタッチしてきましたが、方法論がAIによって変わってきています。

- アーッ!!! Adobe Photoshopの新AI機能が凄いーッ!!!

すると、動画でも可能になってきています。

- Adobe Sensei

## Stable Diffusion

2022/8月に画像生成AI**Stable Diffusion**がリリースされ、大きな変化が起きています。

Stable Diffusion: テキストの入力によりその指示内容の画像を生成することができるAI

これを受け画像だけではなく

- 動画
- 3Dモデル

を生成するシステムもあっという間にリリースされています。興味ある人は調べてみましょう。

## 音声処理

## 音声合成

- 原稿読み上げは音声合成アプリに任せるケースが増えている
- [ゆっくりMovieMaker](#)



## 歌声合成

- VocaloidもAIによりより人との声と判別するのが難しいほどのクオリティになっている。
- 調教(細かいチューニング作業)をAIが行うようになってきてイル
- 美空ひばり VOCALOID:AI™ - AIに関する取り組み

## ディープフェイク

ディープフェイク:本来の意味は「機械学習アルゴリズムのディープラーニングを利用して、2つの写真や動画の一部をスワップ（交換）させる技術」。しかし現在、定義が広がっており、単に「フェイク動画」「偽動画」のことを指すことが多い。

これらによりフェイクニュース(偽情報)が拡散されることとなる。  
一人一人がAIのリテラシーを高めることが重要。

# AIの技術とは

## コンピュータ自ら学習する機械学習

AIはシステムを作るだけでなく、多くのデータによって学習することによって精度が上がっていきます。

これを自動化したのが**機械学習**です。

- MNIST：手書き数字の画像データセット

## 教師あり学習、教師なし学習

学習のもとになるデータによって二つに分かれる。

### 教師あり学習

- 猫の写っている写真
- 猫が写っていない写真

この二つを学習させる方法

### 教師なし学習

多くのデータを特徴によって分類して！という学習方法

- 「グーグルの猫」にみる大量データの重要性 AI開発には質への理解も不可欠

## 強化学習

何度も試行錯誤して、最も良い結果(報酬)を得られた行動を覚えていく手法で、**報酬の最大化**がキーワードとなります。

- 【物理エンジン】 AIにブランコを学習させたら体操選手になった【強化学習】

## 深層学習(ディープラーニング)

ニューラルネットワーク：人間の脳の構造を真似したモデル

ディープラーニング：ニューラルネットワークを4層以上に  
した機械学習の方法

- 【初心者向け】ニューラルネットワークがなんとなくわかる

重みをうまく設定することで、分類などができるようになります。(ここだけで、15回抗議できるので、この辺で...)

## 転移学習

学習したAIがあれば、それを他の分野に転用することができるかもしれない。

うまく利用できれば、

- 大量のデータ
- 学習にかかる時間

これを減らすことができます。