

(YouTube ピクアカ)

データサイエンスとは何か、を理解する

【はじめてのデータサイエンス PART1】

AI・機械学習・データ分析の基本を学ぼう

データサイエンス概論

1. AI・機械学習の重要概念の体系的理解
2. ディープラーニング技術の特性
3. 各種の機械学習モデルの特徴
4. 機械学習の事例紹介

データサイエンスとは ①

- データから有意義な**インサイトを抽出**する学問
- 別々の目的をもつ**複数学術分野の集合体**
- 分野の専門知識、プログラミングスキル、
数学・統計の知識などと**組み合わせ**て使用

【資料】インサイト

インサイトは、直訳すると「洞察」「本質を見抜く」という意味です。マーケティングの分野においては、「消費者の隠れた心理」を表す言葉として一般的に用いられます。

顧客インサイトや消費者インサイトと呼ばれることもあり、顧客自身も気づいていない**動機や本音**を表す言葉です。様々な**購買行動の根底にある要因**がインサイトであり、それらを探り当てることが商品開発やサービスの提供において重視されています。

データサイエンス

統計学

多変量解析

機械学習

データマイニング

人工知能

その他

【資料】データマイニング

収集された情報のなかから**傾向や関連性**を見出す
分析手法。

売上予測や市場動向など、ある事象の**発生予測**や
データの類似性から**新たな仮説を立てる**など、膨大なデータを**ビジネスに活かす**のに役立つ。

データサイエンスとは ②

シンプルに言えば

「**個々**の観測対象から、**集団の性質を抽出**する技術」

データサイエンティストに必要な**素養**

1. コンピュータ技術
2. 統計学・数学・機械学習
3. ビジネス能力（データ分析結果をビジネスに活用）

<一般に>

ビジネス寄りの「**業務系**」

エンジニア寄りの「**実装重視系**」に『峻別する』

身近に溢れる機械学習技術の社会実装

自動運転、音声認識、検索エンジン、
レコメンド、医療診断 等

- 機械学習は作り上げた学問ではない
- 手法や理論は、
統計学に基づいた
従来の分析手法の延長線上に発展

なぜ近年になって注目？

機械学習の進歩にある背景

- 燃料：ハードウェアの性能向上
- 材料：ビッグデータの貯蓄
- 賢さ：アルゴリズムの改善

オープンソースのプログラミング言語

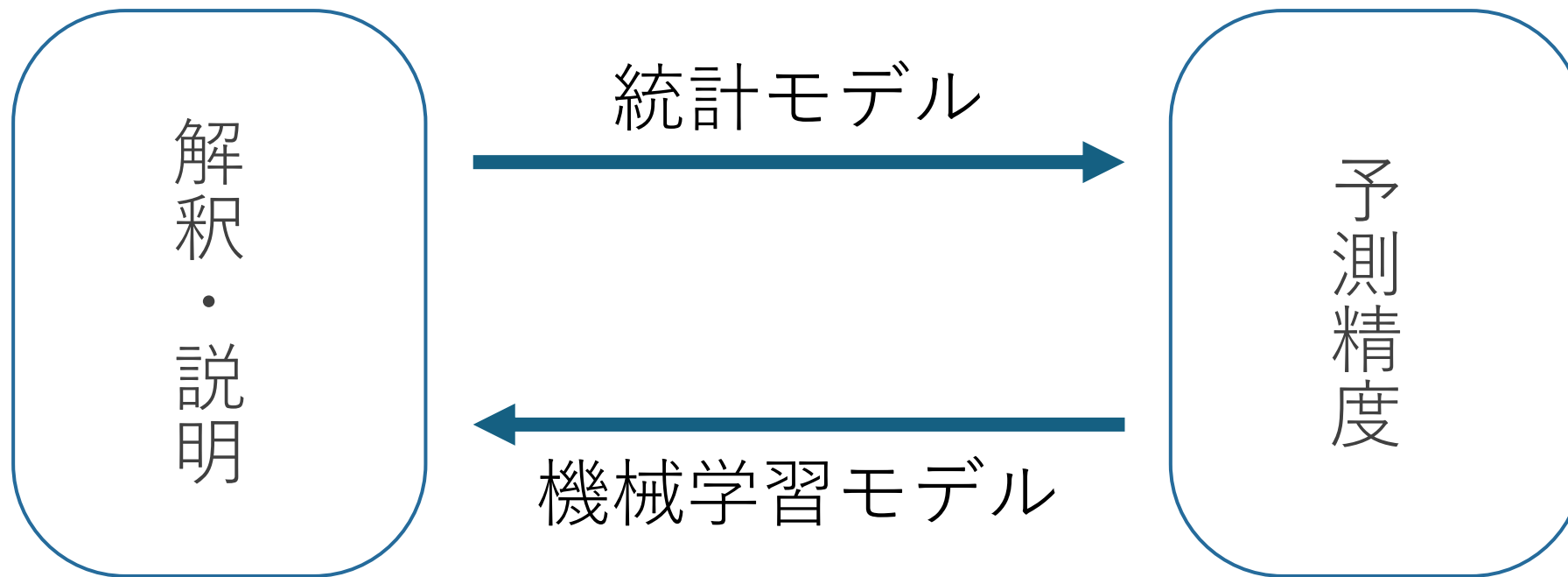
Python、Rなど

【資料】 R言語

統計解析のために開発、他の分野には不向き
活用するためには統計学に関する知識が不可欠
ソースコードは簡単
ベクトル処理・行列演算、
データの可視化が得意

機械学習 VS 従来の手法（統計モデル）

違いは？



機械学習

ビジネスの場面に適している



スピードや柔軟性が要求される

機械学習

高精度の「予測モデル」をつくること

学習と予測

学習：蓄積したデータに基づき、
機械が法則やパターンを見出す

予測：学習済みモデルに未知データを入力
予測値を算出

人工知能と機械学習の位置付け 23'27"

人工知能：

知的な機械、

特に、知的なコンピュータプログラムを作るための科学技術

ジョン・マッカーシー

機械学習：

明示的にプログラムしなくても

学習する能力をコンピュータに与える研究分野

アーサー・サミュエル

二つの人工知能

ルールベース：

ルールに従って判断

機械学習：

学習データから自動的に法則を見つける

機械学習の「学習」と「予測」

学習：データからパターンを見出す

予測：学習済みモデルに未知データを入力し
予測値を算出

機械学習の主な領域

教師あり学習：クラス分類、回帰（数値予測）

教師なし学習：クラスタリング、次元圧縮

強化学習：アルゴリズムはバンディット、Q学習

教師あり学習

① クラス分類

カテゴリ分けする境界線を見出す

② 回帰

できるだけ多くの数値を表現できる
関数を見出す