

エピローグのストーリー

- 清原は九条とともに、地方自治体がかかえる問題に対して機械学習を用いたソリューションを提供する会社を立ち上げる
- そこに就職先のビジョンに失望したさやかが合流する

機械学習エンジニア育成のために



p.204
1～4コマ目

機械学習技術者に求められるもの

- 理論に強いエンジニア
 - 問題に適した機械学習手法を選択できる
 - 新しい技術をチームに導入できる
- 実践に強いエンジニア
 - 対象業務を深く理解し、データの整備ができる
 - 実運用に際して、性能の見極めができる
- 実装に強いエンジニア
 - データの取得・管理、運用環境が構築できる

理論に通じている技術者

p.111 2コマ目

ん〜…

特徴抽出も学習の対象にできる点かな
これまでは音声や画像の認識には複雑な手順に
基づいた特徴抽出処理が必要だったんだ
でもディープニューラルネットワークでは

画像データや音声信号をそのまま入力して
単純な特徴表現から段階的に複雑な構造を
取り出すことができているよ
とても高い性能を実現しているよ

ディープラーニングが得意なもの

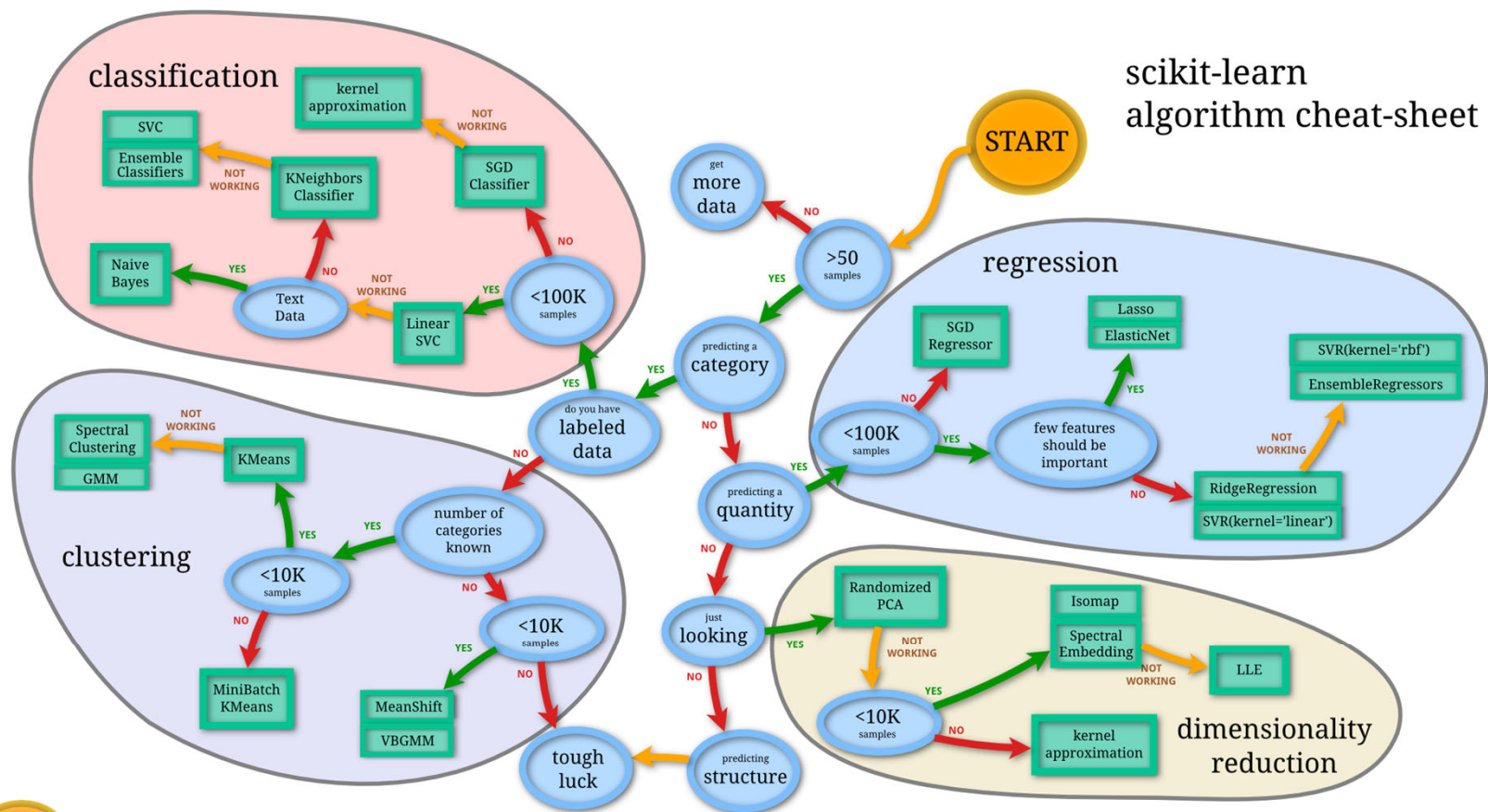
- 音声認識
- 画像認識
- 自然言語処理



なるほど

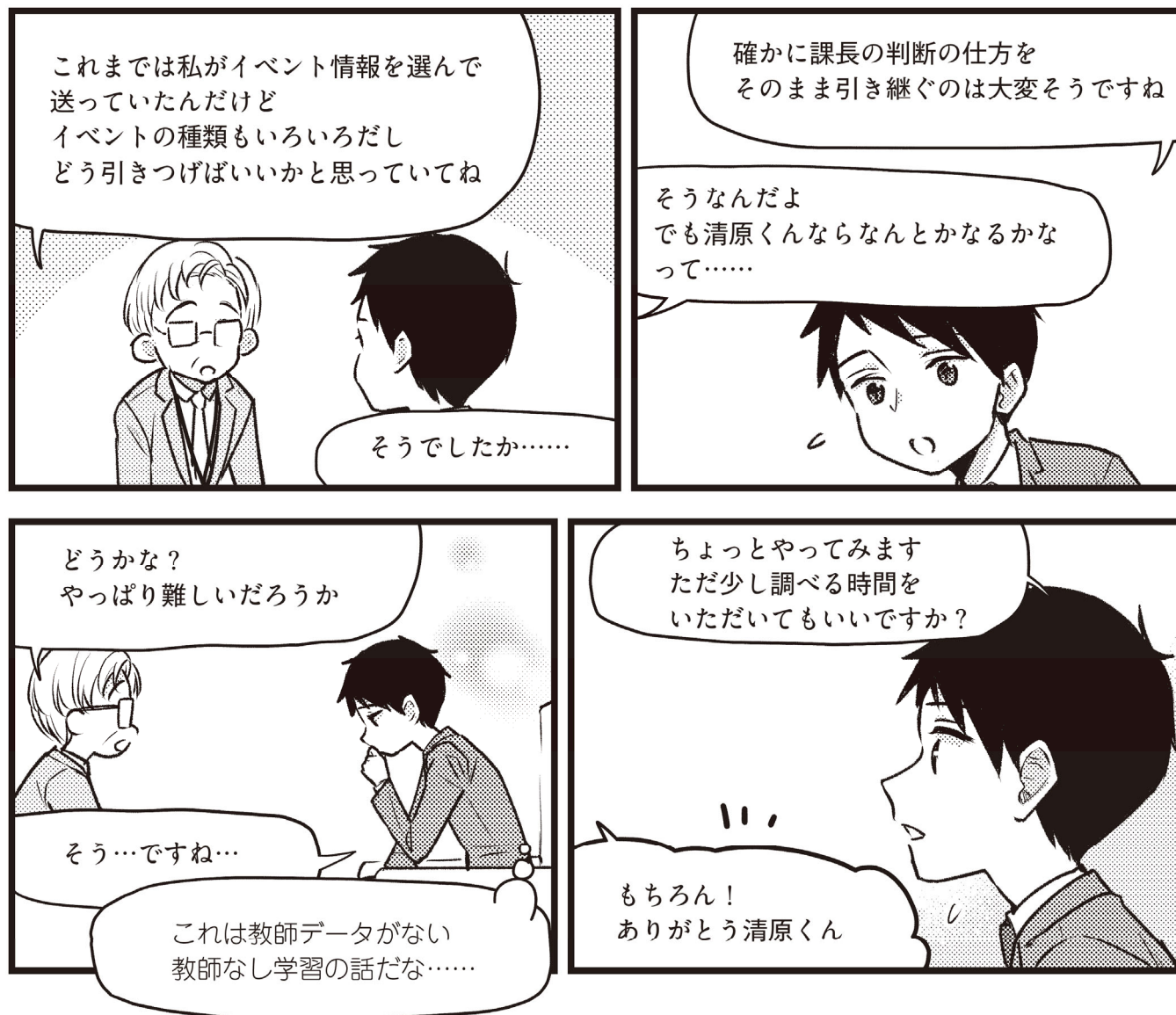
理論に通じている技術者

- 問題を正確に定義し、適用する手法を見極める



実践に携わる技術者

p.167 4-7コマ目



実践に携わる技術者の育成

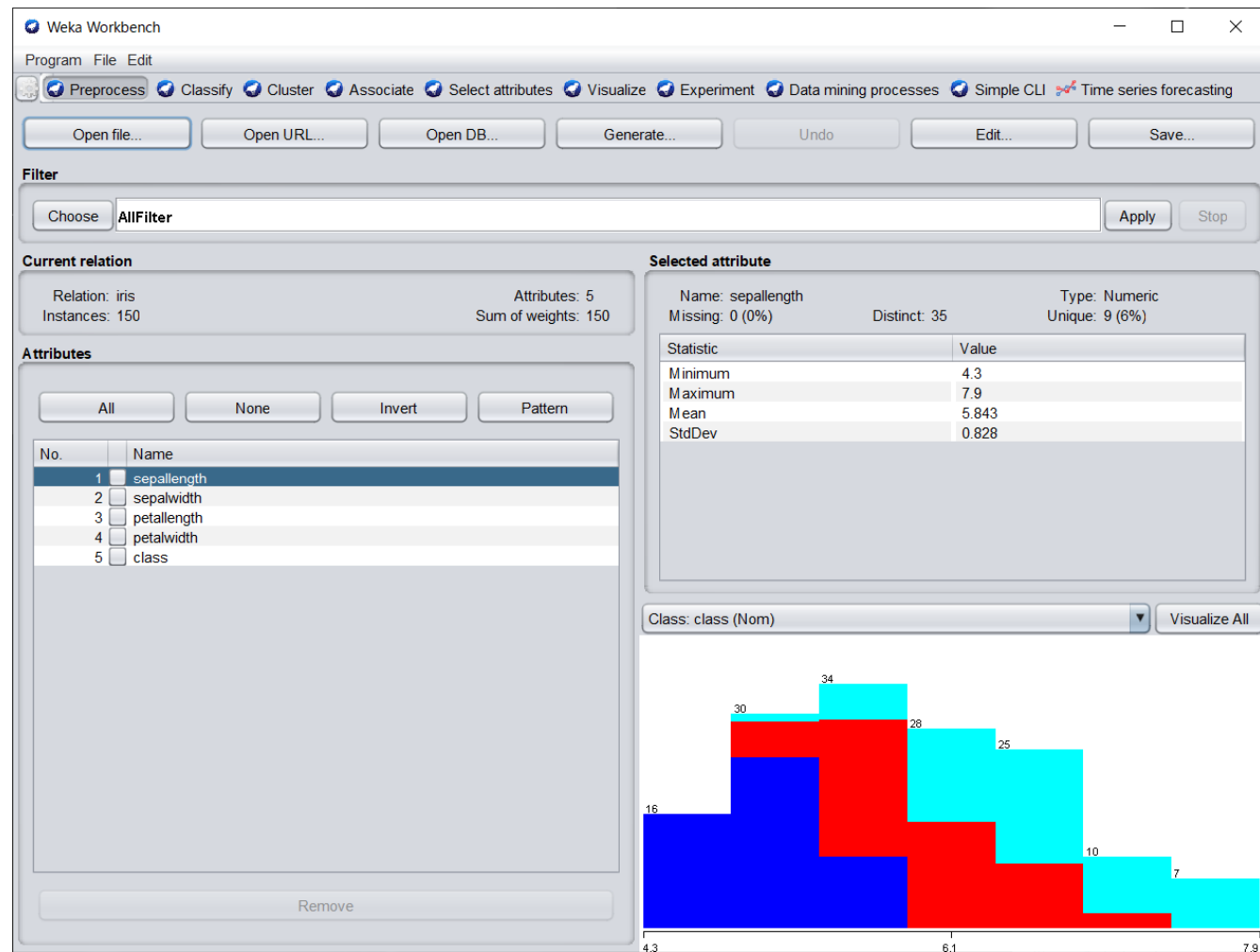
- 開発経験
 - どの程度の性能でどの程度の効果をもたらすか判断
 - 機械学習以外の選択肢も持つ
- データの前処理
 - 生データから、ツールへの入力に変換するまでが大変
 - 欠損値・異常値の扱い、正解の信頼性の検討が必要
- 結果の妥当性の判断
 - データの性質に基づく正しい評価法の選択

開発を支えるツール(1)

- GUIツールWeka
 - Waikato Environment for Knowledge Analysis
 - 機械学習のアルゴリズムを実装したJavaライブラリ
- データファイルを直接操作できるGUIを持つ
- ライセンスは GNU GPL
 - プログラムの実行・改変・再配布が自由
 - ただし二次的著作物に対してもGNU GPLが適用される

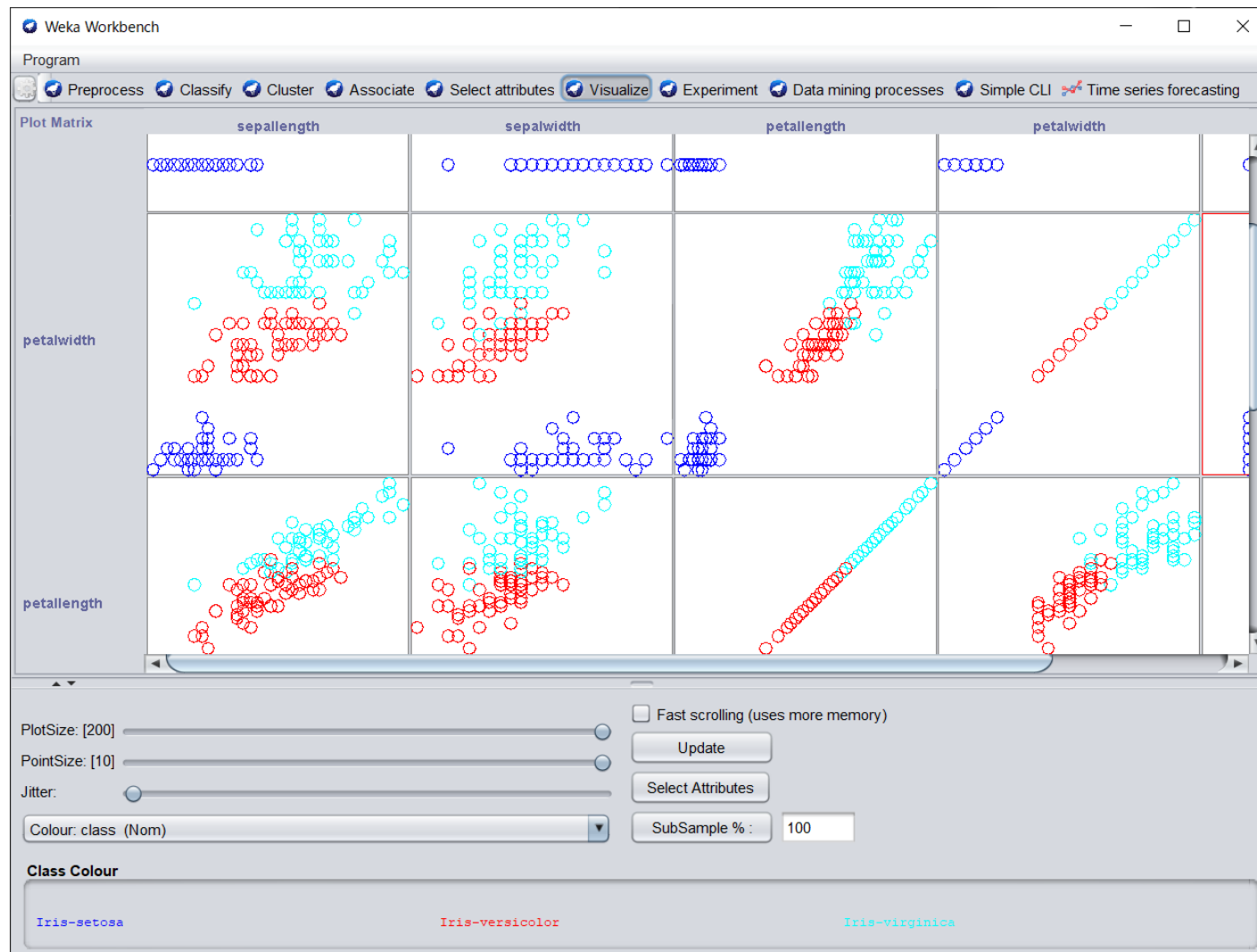
開発を支えるツール(1)

- データの読み込み、分析画面



開発を支えるツール(1)

- データ可視化画面



開発を支えるツール(1)

- 識別の実行画面

The screenshot displays the Weka Workbench interface with the Classifier tab selected. The classifier chosen is J48 -C 0.25 -M 2. The Test options are set to Cross-validation with 10 folds. The Classifier output shows stratified cross-validation results and a detailed accuracy by class table.

Classifier output

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      144           96
Incorrectly Classified Instances     6             4
Kappa statistic                     0.94
Mean absolute error                  0.035
Root mean squared error              0.1586
Relative absolute error              7.8705 %
Root relative squared error          33.6353 %
Total Number of Instances           150

=== Detailed Accuracy By Class ===
```

	TP Rate	FP Rate	Precision	Recall	F-Meas
Iris-setosa	0.980	0.000	1.000	0.980	0.990
Iris-versicolor	0.940	0.030	0.940	0.940	0.940
Iris-virginica	0.960	0.030	0.941	0.960	0.950
Weighted Avg.	0.960	0.020	0.960	0.960	0.960

```
=== Confusion Matrix ===

 a b c  <-- classified as
49 1 0 | a = Iris-setosa
 0 47 3 | b = Iris-versicolor
 0 2 48 | c = Iris-virginica
```

Overlaid on the Weka window is a separate window titled "Tree for 22:04:46 - trees.J48" showing the decision tree structure:

```
graph TD
    A(petalwidth) -- "<= 0.6" --> B[Iris-setosa 50.0]
    A -- "> 0.6" --> C(petalwidth)
    C -- "<= 1.7" --> D(petalwidth)
    C -- "> 1.7" --> E[Iris-virginica 46.0/1.0]
    D -- "<= 4.9" --> F[Iris-versicolor 48.0/1.0]
    D -- "> 4.9" --> G(petalwidth)
    G -- "<= 1.5" --> H[Iris-virginica 3.0]
    G -- "> 1.5" --> I[Iris-versicolor 3.0/1.0]
```

開発を支えるツール(2)

- Jupyter notebook
 - ブラウザで実行できるPython開発環境

準備

必要なライブラリ等を読み込みます。

```
In [1]: import tensorflow as tf
        from tensorflow import keras
        import numpy as np
        import matplotlib.pyplot as plt
```

データの読み込み

```
In [2]: fashion_mnist = keras.datasets.fashion_mnist
        (train_images, train_labels), (test_images, test_labels) = fashion_mnist.load_data()
```

データサイズの確認

```
In [3]: train_images.shape
```

```
Out[3]: (60000, 28, 28)
```

```
In [4]: test_images.shape
```

```
Out[4]: (10000, 28, 28)
```

開発を支えるツール(2)

- Google Colaboratory
 - クラウドで実行可能なJupyter notebook
 - 機械学習関係のライブラリはインストール済み

実装に強い技術者

p.72 2コマ目

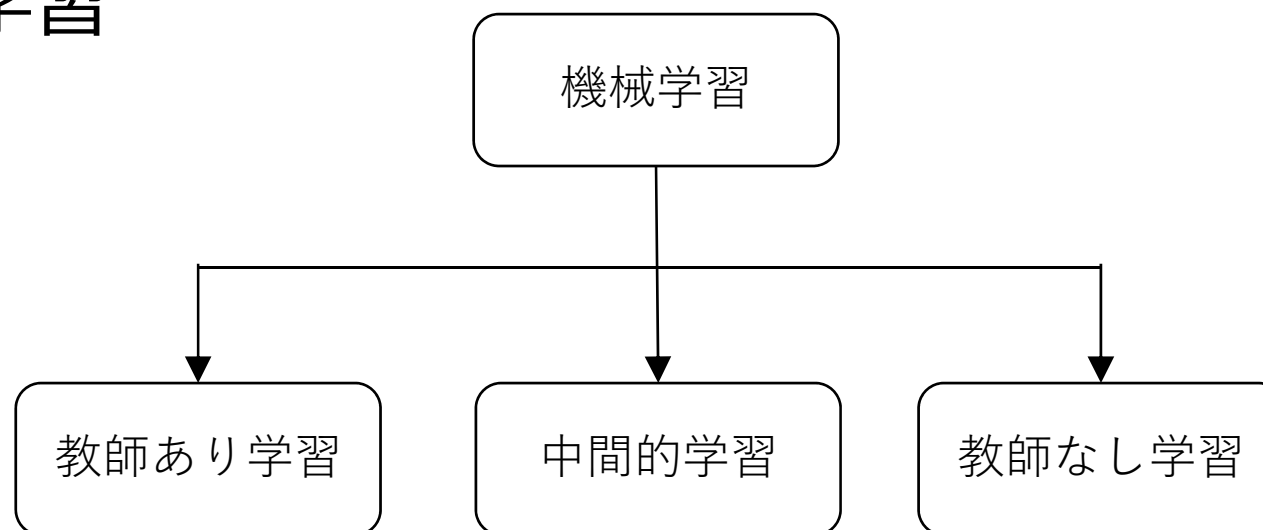


実装に強い技術者

- 社内データの有効活用
 - 実は活用可能なデータは社内に眠っている？
 - 適切なアクセス権限を設定し、データを活用できる人を増やす
 - 社外流出防止等のセキュリティ強化も重要
- 快適な開発環境
 - 機械学習に用いるデータは一般に大規模
 - 高速なアクセスができる環境構築が重要

事例紹介 - 強化学習の導入を通じて -

• 強化学習



正解付きデータ

年齢	性別	時刻	購入
35	男	16	Yes
24	男	9	Yes
22	女	21	No



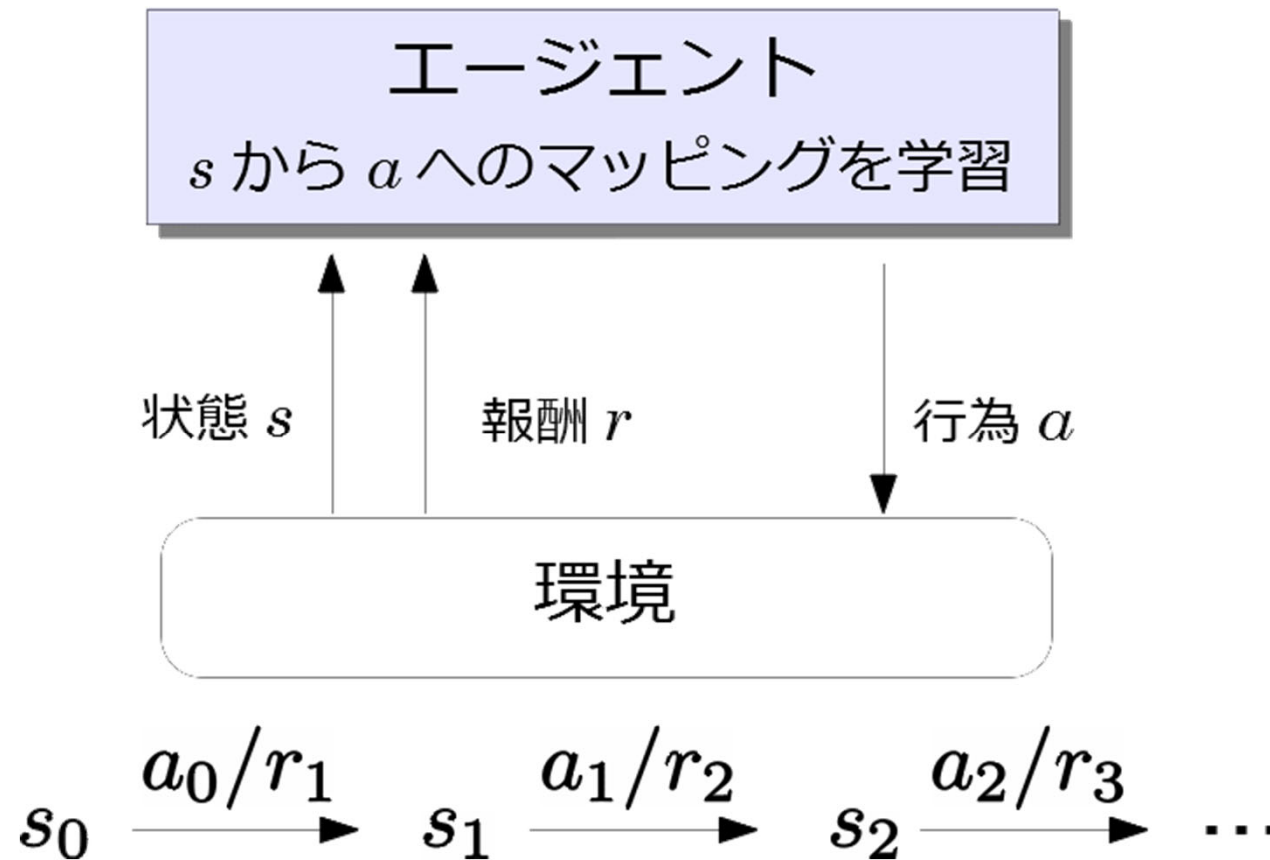
強化学習

正解なしデータ

長さ	幅	重さ
15	6	16
24	8	19
32	7	18

強化学習

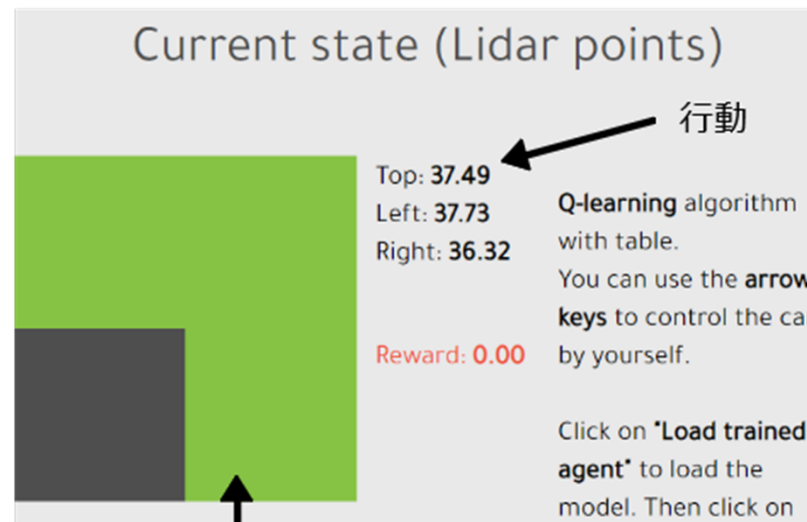
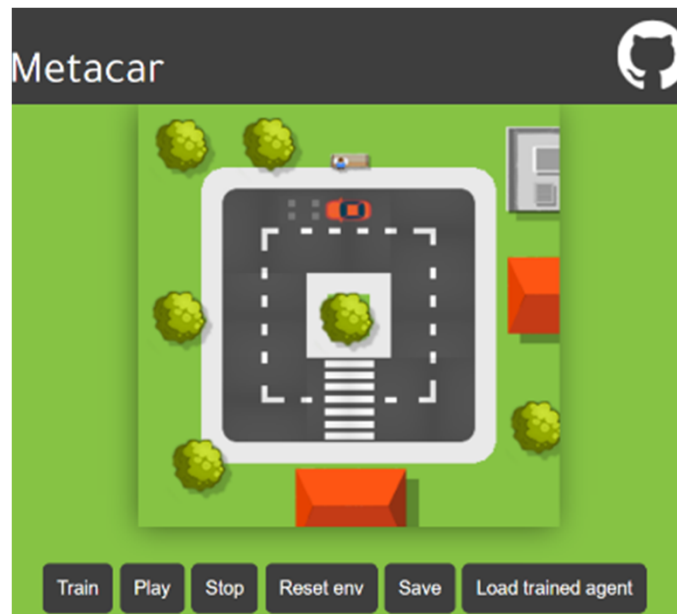
- 環境の中で行動するエージェントの学習
 - 報酬が遅れて与えられる



強化学習

- 強化学習の例：Metacar

<https://www.metacar-project.com/qtable.html>



Top:前進
Left: 左
Right: 右

車の前方の状態

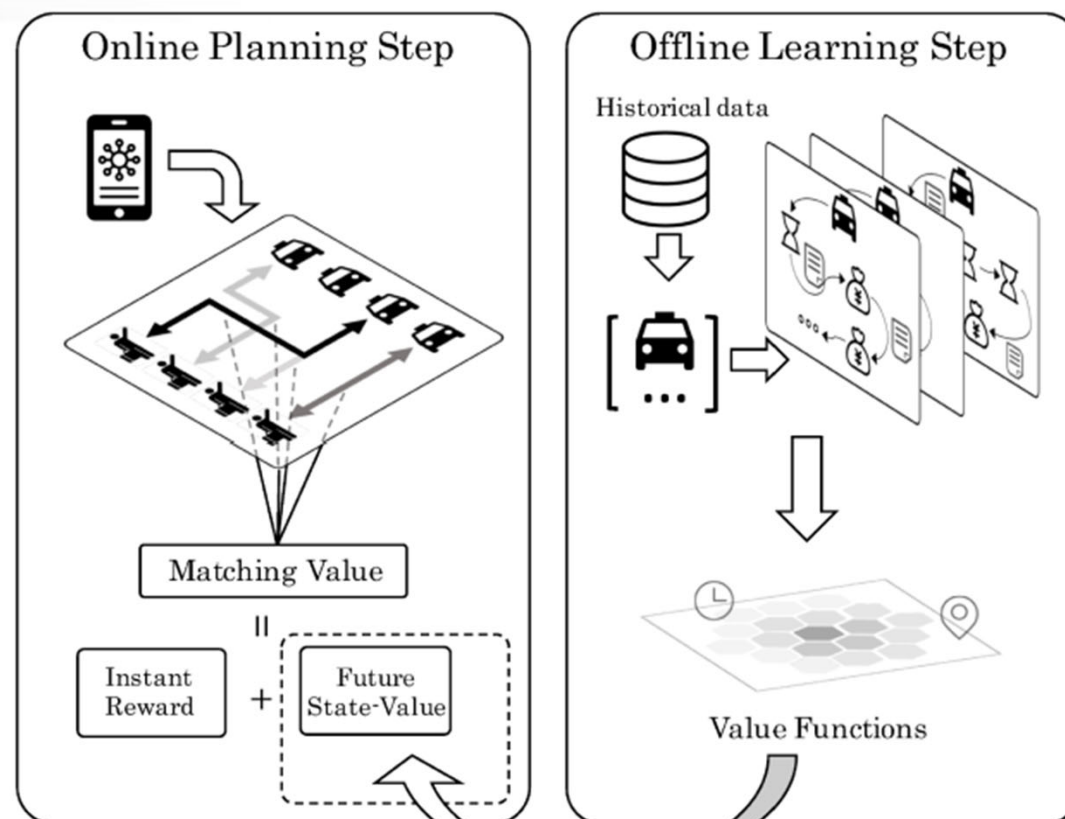
1. Load trained agentでモデルをロード
2. Playで車をスタート
3. 左右矢印で走行を妨害

強化学習を用いた事例紹介(1)

- 滞納債務の取り立て（IBM）
 - 債務者に対するアプローチ（督促状、電話など）の手順を強化学習の枠組で学習
 - 例）状態：電話催促済み、滞納額1万ドル、期間6ヶ月
行為：訪問
 - ニューヨーク州の徴税部門で2009年12月から稼働しており、3年間で100万ドル以上の税収増の効果が
あると見積もられている

強化学習を用いた事例紹介(2)

- タクシー配車サービスDidiでの配車最適化



過去のデータから
時間・地域の組合せ
を状態として価値
関数を学習

「即時報酬+将来の期待収益」の最大化

強化学習を用いた事例紹介(2)

- 配車最適化モデルの転移学習
 - 一つの都市で学習したモデルを他の都市に適用
 - 価値関数をディープニューラルネットワークで近似
 - 時間・地域に新たな特徴（需要、供給等）を加える
- 参考資料
 - <https://speakerdeck.com/pacocat/reinforcement-learning-applications-in-taxi-dispatching-and-repositioning-domain>

まとめ

- よいチームを作る
 - 理論・実践・実装のすべてを完全にカバーできる人はいない
 - 上記3つのいずれかに軸足を置き、他の2つは何をやっているかがわかる程度になっておく
- 必ず一定時間を勉強に使えるように
 - 技術が変化するので、継続的な勉強が必要
 - 社外、大学も巻き込んだコミュニティへの参加