

ローカルAIマップ例 機械学習（ML）マップ

電力中央研究所 エネルギーイノベーション創発センター
デジタルトランスフォーメーションユニット
ユニットリーダー 堤 富士雄

2019年6月6日 JSAI2019 企画セッション AIマップ

AIマップβは全体像なので、
個々の分野のことは分からない。
もっと知りたい人に向けたガイドが必要。

良いガイドが増えれば増えるほど、
新人が増えて分野が活性化する。

MLマップの背景と目的

背景：

- 機械学習（ML）が高度化、利用が簡単に
→活用事例が増加、事業活動に必須化
- ユーザが、ML活用のための道案内となる資料は少なかった
ビジネス寄りの大雑把な解説 OR
専門性が高く難解な資料かの両極端
- 異分野研究者や現場技術者がML活用に役立てるマップが必要

目的：

- ML活用する研究者や技術者が、MLで何ができて、何ができないのか、どうすれば活用できるのか、利用面からの全体像を把握できるMLマップを作成
- 各MLの位置付けや、活用の前提となる基礎知識を得るのに役立つ資料



秋庭伸也、杉山阿聖、寺田学、監修：加藤公一「見て試してわかる機械学習アルゴリズムの仕組み 機械学習図鑑」翔泳社、2019年

良書。すばらしい。

MLマップの目指すところ

MLは色々できるようだけど、私の仕事（研究）で、どう使えばいいんだろう？



MLが専門ではない研究者
電力などの現場技術者

現在のMLマップ



そうか。ここに画像認識のMLを組み込めば、効率上がるぞ。そのためには〇〇のデータが必要だな。

未来のMLマップ



そうか。今はまだ、この問題は解けないけど、〇〇データを準備しておけば、いずれML活用できるな。

MLの使えない技術者



ある程度の基礎知識
統計・プログラミング

1. 活用事例・研究事例のインプット

2. 各ML手法の得手・不得手の把握

3. 適用に必要な準備・仕組みの理解

4. 実際に使えるMLツール・サービス・サンプル

MLマップが対象とする範囲

業務や研究の現場で問題を見つける方法
変えられない構造の発見方法

現状の技術水準と今後の発展をにらんで、ターゲットとすべき部分問題、MVCの設定

改善・改良ステップの仕組みを考案

どちらかというとIoT的なノウハウ

MLの使える技術者



ML適用すべき問題を発見・定義し、ユースケース・サービスモデルを考案し、IT企業等と連携して実行・システム構築できる。

MLマップの構成案

1. 活用事例・研究事例のインプット

事例マップ

難度マップ

2. 適用に必要な準備・仕組みの理解

準備と利用

処理の流れと各手法の位置づけ

3. 各ML手法の得手・不得手の把握

問題のタイプ

MLの変遷と手法群の特徴

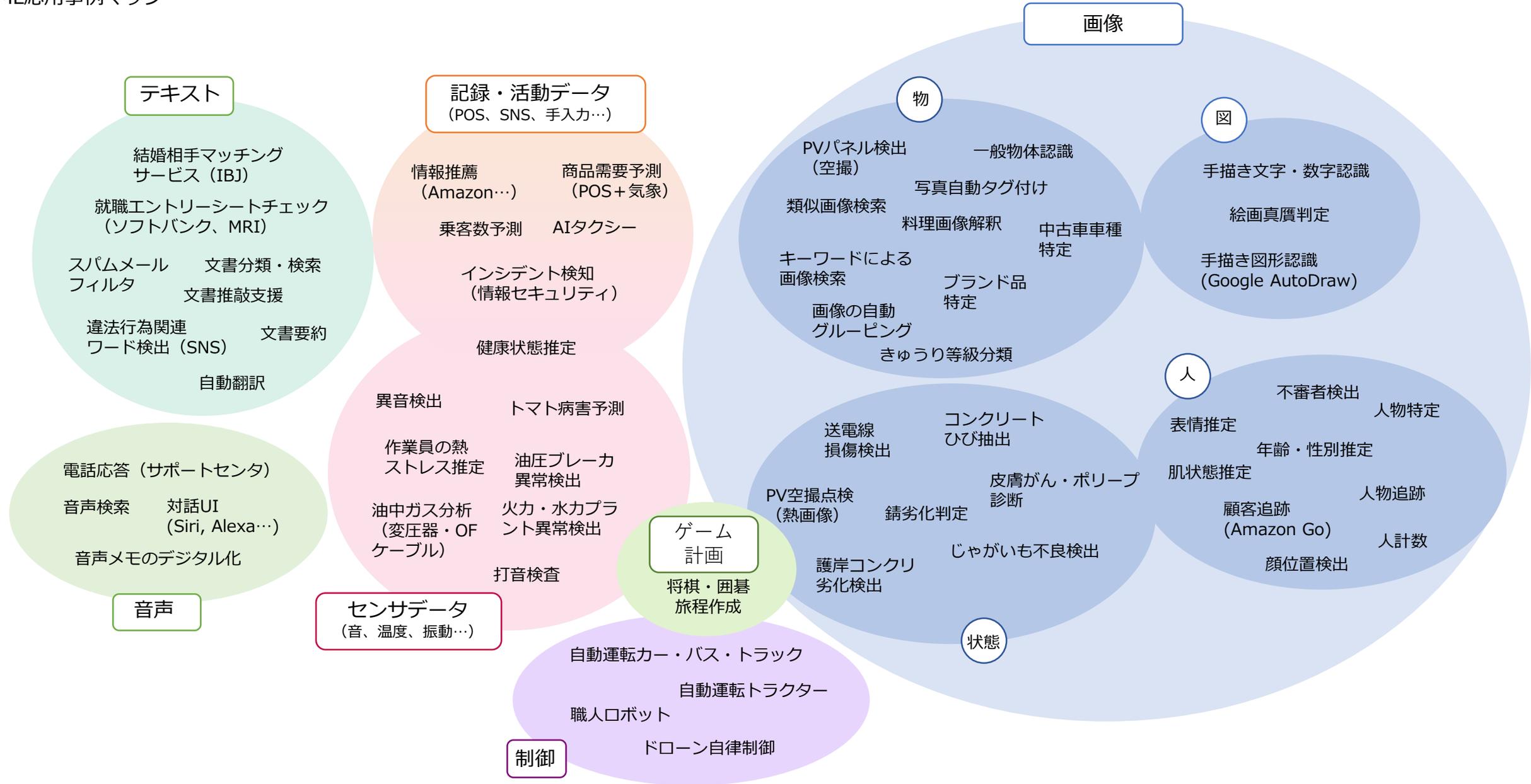
苦手な分野と今後の発展

4. 使えるMLツール・サービス

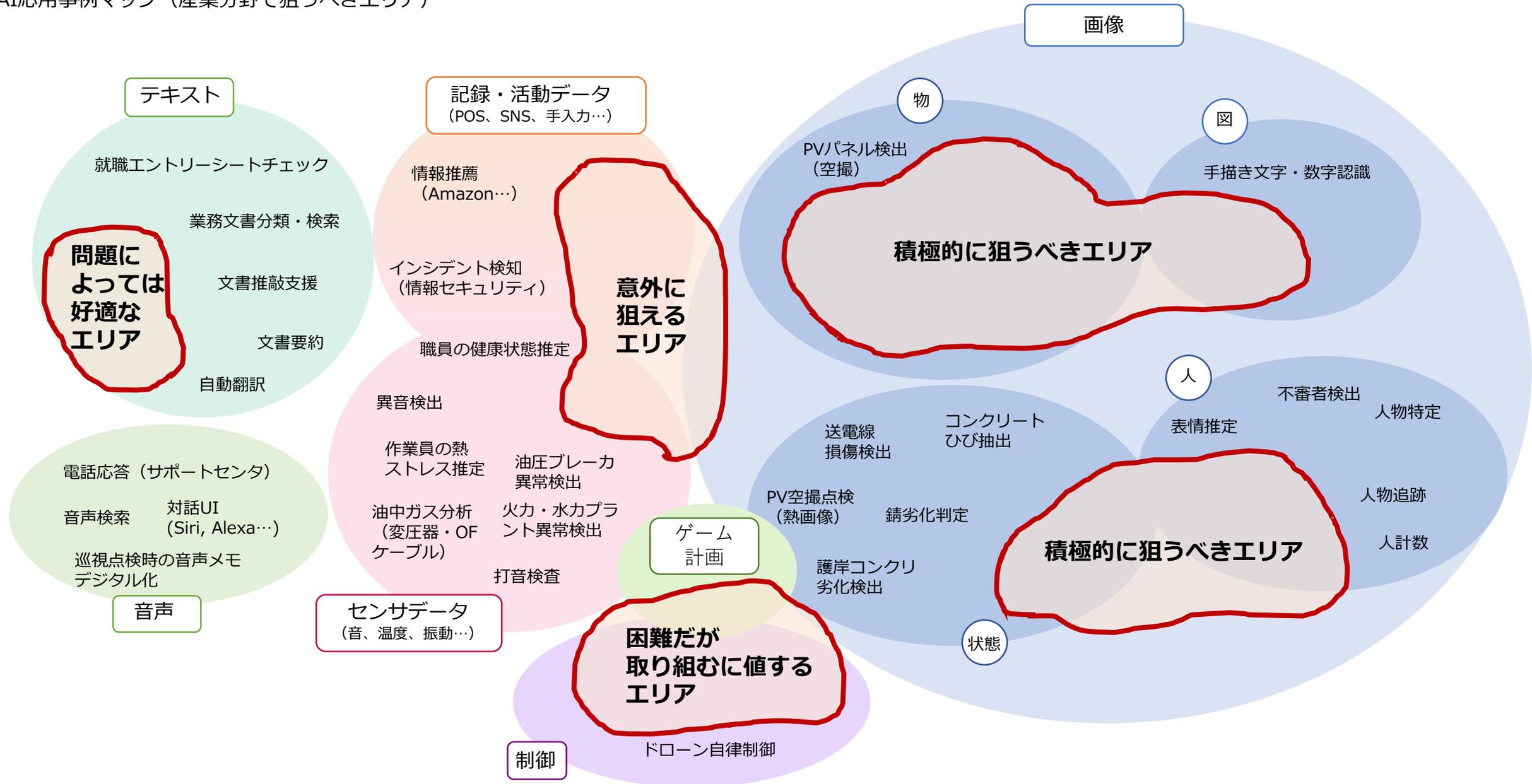
手法選択のチートシート

ML活用環境と連携したチュートリアル

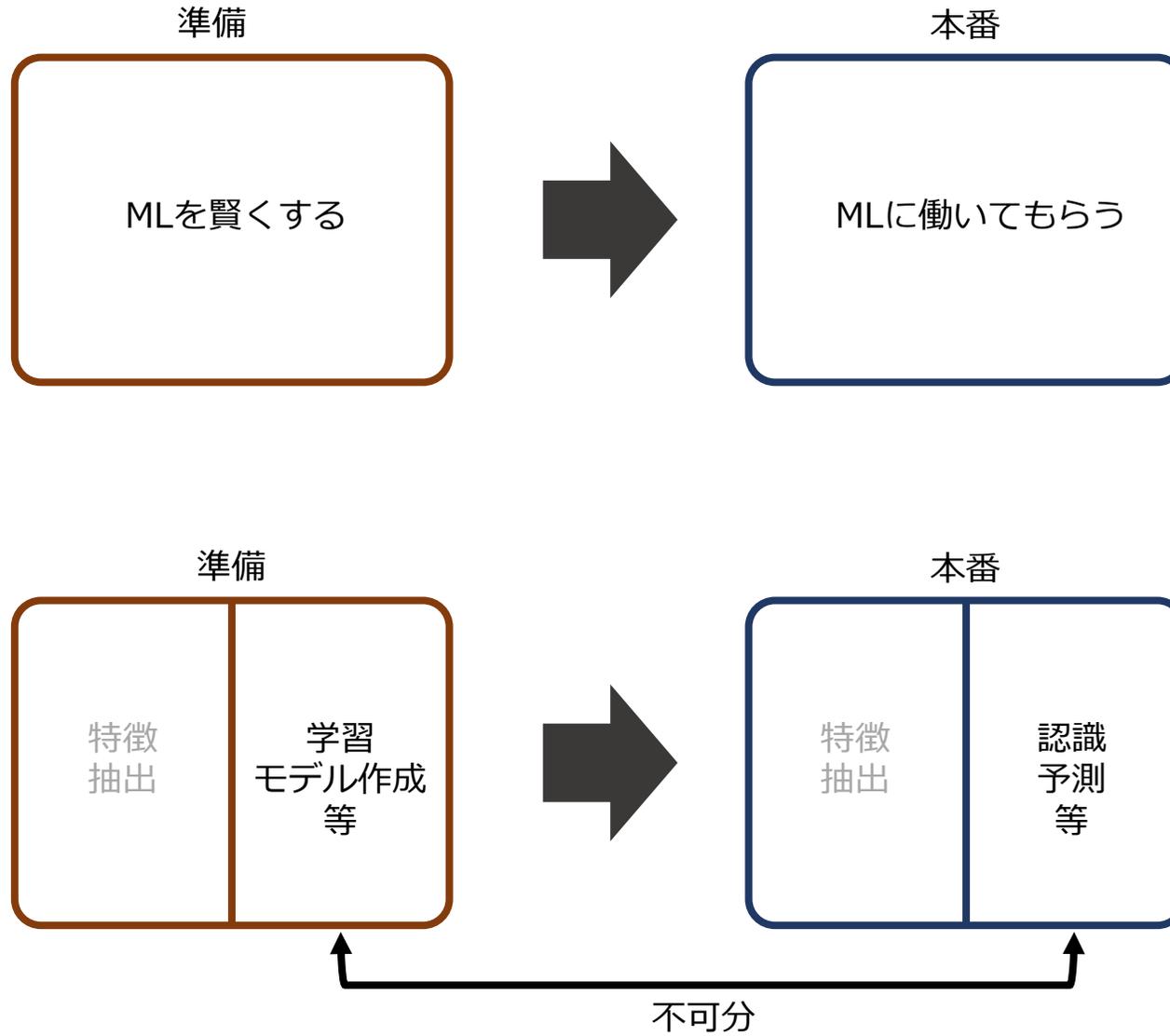
ML応用事例マップ

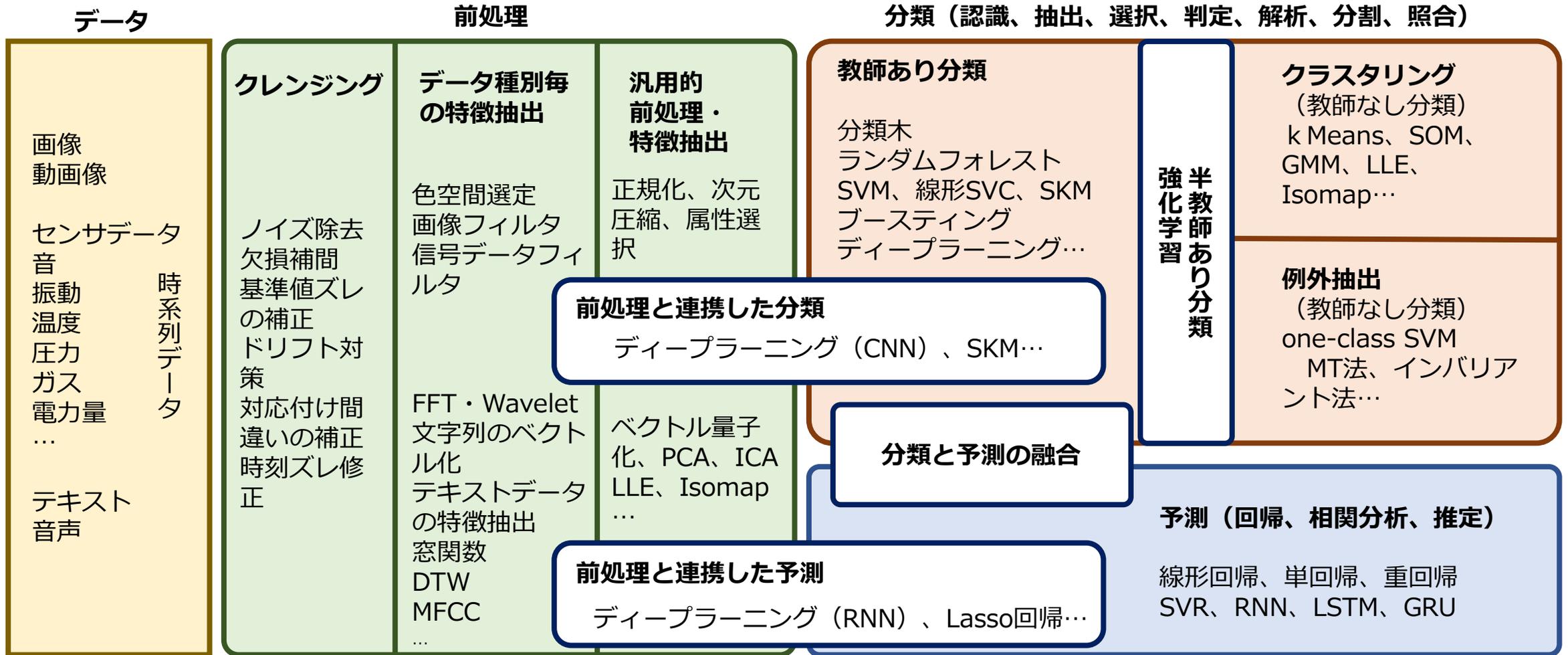


AI応用事例マップ（産業分野で狙うべきエリア）

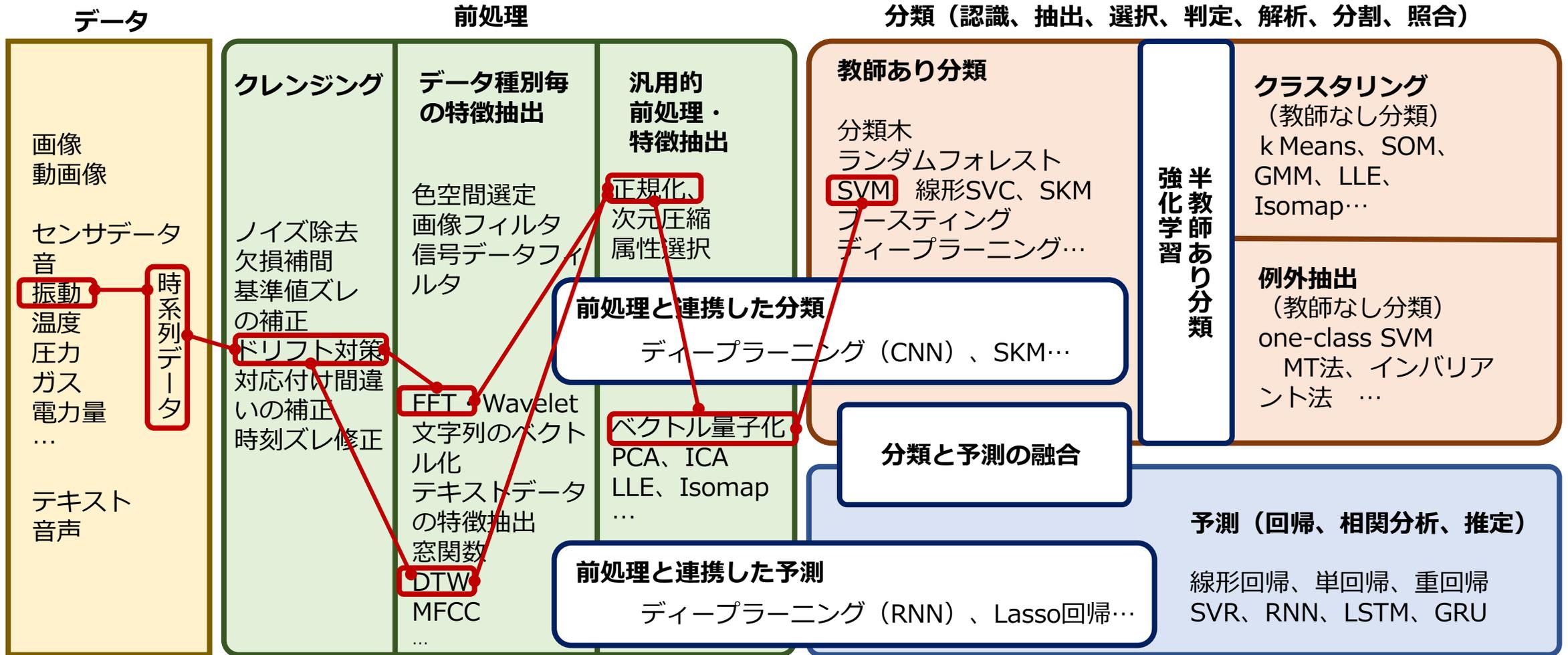


準備と利用

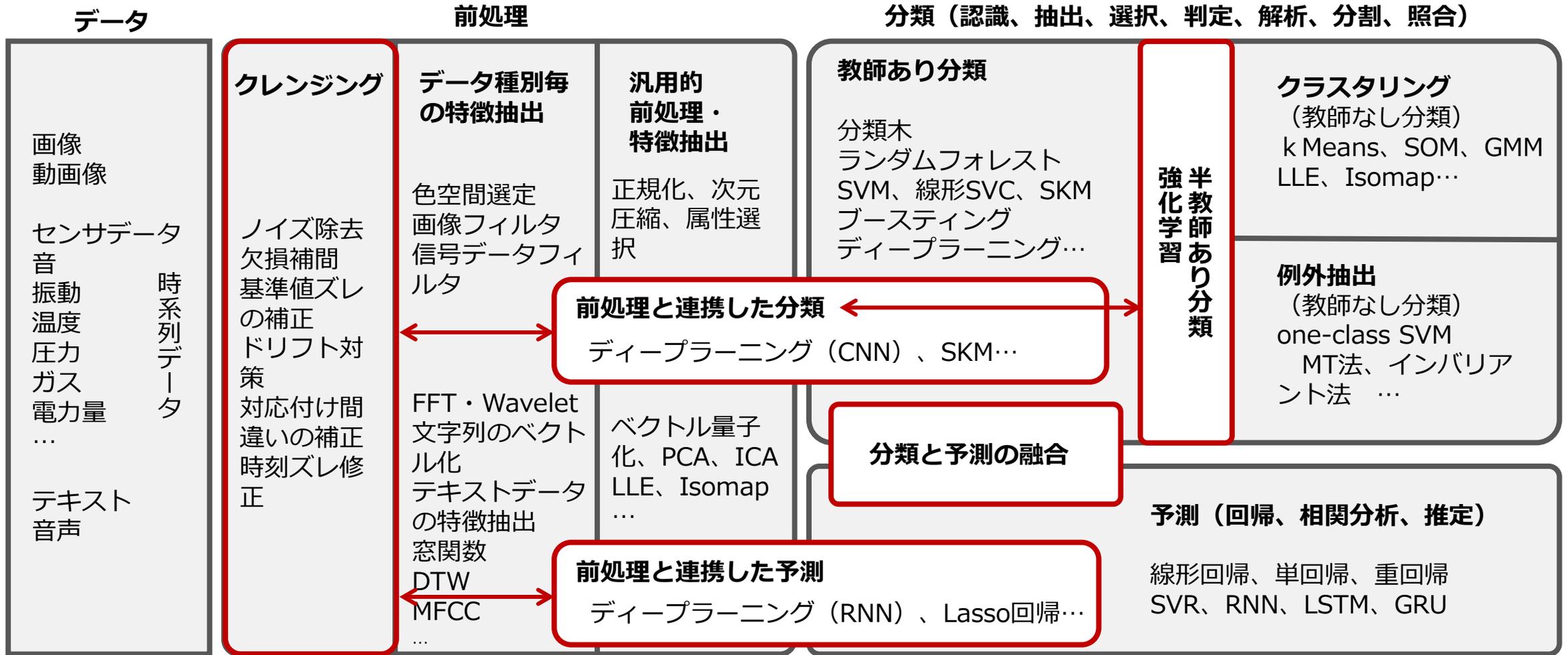




処理の流れと各手法の位置づけ



振動時系列データの異常診断の例



今後注力すべき分野

アンサンブル

- 複数の組み合わせで性能を上げる手法
- ○ 単体よりもよくなることもある
- × 単体よりも計算量が多い
- × 単体の場合よりも判定根拠が分かりづらい

分類木

- 条件分岐で判定する手法
- 2000年以降にアンサンブルと組み合わせて発展
- ○ 比較的少量のデータでも、それなりに良好な結果
- ○ 計算量が少ない
- ○ 寄与度など、根拠が比較的わかりやすい
- × パラメータ調整（構造決定、連続値の扱い）に手間
- × 回帰の場合、結果が階段状・滑らかでない

分布系／パラメトリック

- データ分布を推定・仮定して判断する手法
- 古くから使われている基本
- ○ 分布が当たれば少ないデータで良好な結果
- × 分布を事前に推定するのが難しい
- × 不明な分布を推定するにはデータ量が必要

ニューラルネットワーク／グラフィカルモデル

- 神経細胞をヒントに作られた手法
- 処理ノードの組み合わせでモデルを表現する
- ○ 近年、ディープラーニングで高精度化
- ○ 特徴量の自動抽出ができるようになってきた
- ○ 因果関係・シーケンスを扱いやすい
- × 多量のデータと計算量が必要
- × 判定根拠は分からない
- × 理論化・体系化が不十分で職人芸が必要
- × センシティブでチューニングが難しい

SVM

- データ群の境界に注目した手法
- ○ 少数の、偏りあるデータでも良好な結果
- ○ 複雑な分類や予測も非線形カーネルで対処
- ○ 万能選手、使い勝手よし
- × 非線形を扱うには計算量が増大
- × データ量の量に対して敏感に計算量が増える

MLの進展マップ

1980以前 1980 1990 2000 2010

アンサンブル

組み合わせによる分類木手法の高度化

多数決により解を安定して出せるようになった

分類木

データ主義・単純な分布

単純な分布に従うデータがあればなんとかなる世界

ベイズ理論の活用
確率分布のデータによる更新

計算手法が高度化し、複雑な現象のモデル化が可能に

分布系
パラメトリック

グラフィカルモデル

ニューラルネットワークの実用化と
限界の露呈

神経細胞をモデルとした機械学習の枠組みの走り
モデルをデータから構成

ディープラーニング

他分野の様々な工夫を取り入れて、
ニューラルネットワークがリバイバル
データの特徴を自動抽出できる場合
が出てきた

ニューラル
ネットワーク

SVM

SVMの実用化
データ群の境界への着目

比較的少ないデータでも高精度な学習が可能に

1980以前

1980

1990

2000

2010

クラシック

適用分野は狭い。
 扱えるデータ量も少ない。
 天才たちが、基本的な原理を
 紙と鉛筆で発明。

- ・ 距離・ノルム
- ・ 正規分布、その他の分布
- ・ 回帰分析
- ・ 分類木
- ・ モンテカルロ法
- ・ パーセプトロン
- ・ 線形SVM

勃興期

適用範囲が急速に増した。
 大量のデータを扱えるよう
 になった。
 役者が揃った。

- ・ ニューラルネットワーク
- ・ ベイズ推定
- ・ アンサンブル学習
- ・ 非線形SVM

計算機の性能向上、さらに求解
 手法が高度化し、難しい問題も
 解けるようになった。

民主化

Google, MS, Amazonらが巨
 額の投資をしオープンソース化
 とクラウド化を推進。
 特徴抽出などの職人芸が徐々に
 不要に。
 高額なシステム構築も不要、
 使い勝手が良くなり、
 誰もが使える便利で安い道具に
 なってきた。
 GPUを使った超並列計算が安
 価・容易に。

MLのExcel化

使えるツールのレベル

どのあたりで使うのか？

レトルト食品レベル：レンジで食べられる。自由は効かない。



学習済みのレディーメイドなデータ分析サービス

Watson, Amazon Recognition/LEX, Google Cloud Speech API...

家庭料理レベル：ちょっとした料理技能が必要だが、失敗は少ない。まあ、食える。



データを準備し、多少カスタマイズして使うサービス

scikit-Learn, Bluemix, Amazon ML, Google Cloud ML, ...

料理人レベル：良い道具が揃ってきたが、基本はプロ仕様。苦勞した拳句、不味い料理が簡単にできる。



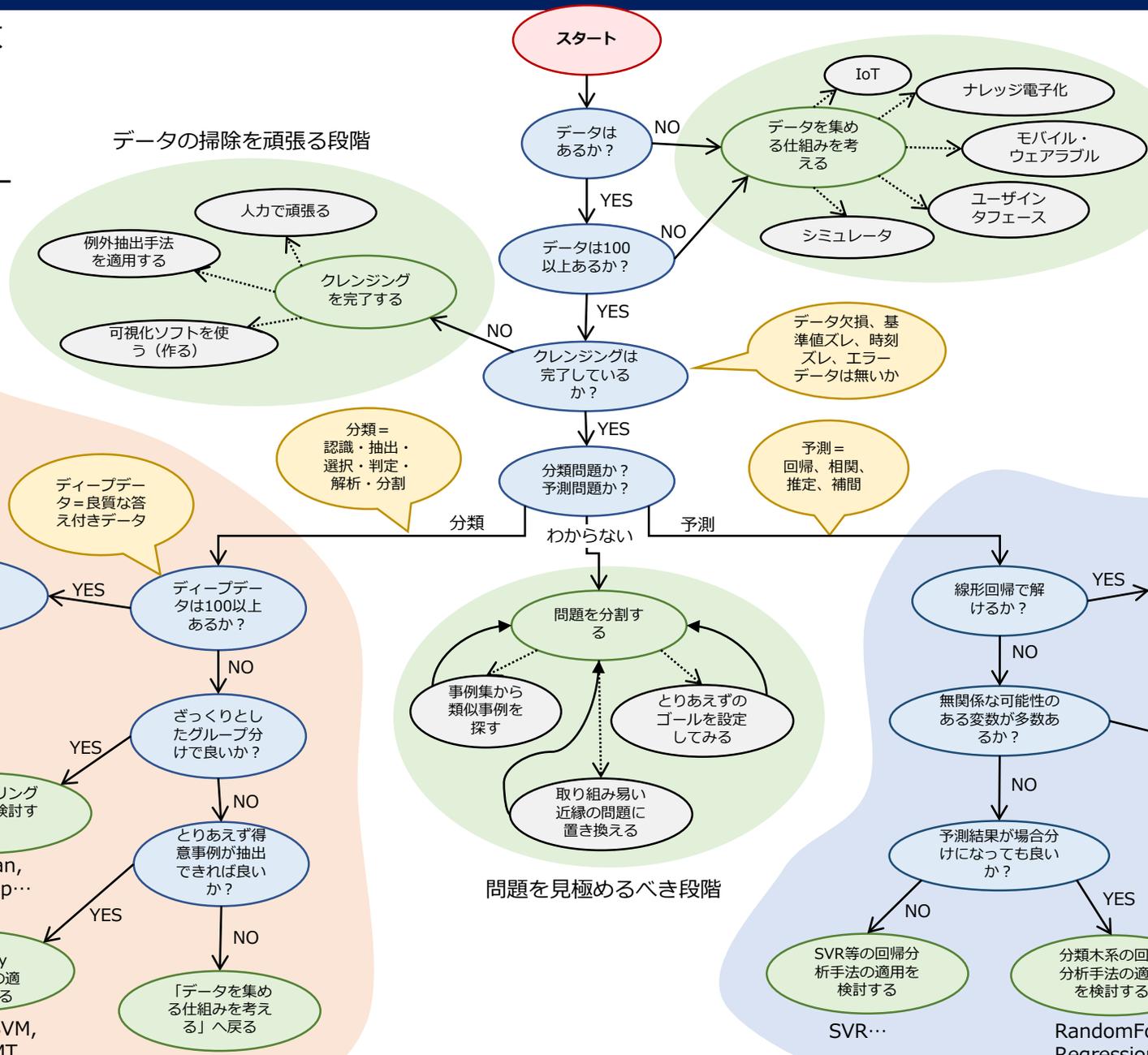
ライブラリを利用しフルスクラッチで作成

MXNET, TensorFlow, CNTK, ...

下のレベルを使いこなすほど大きな**差別化**

手法選択シート改

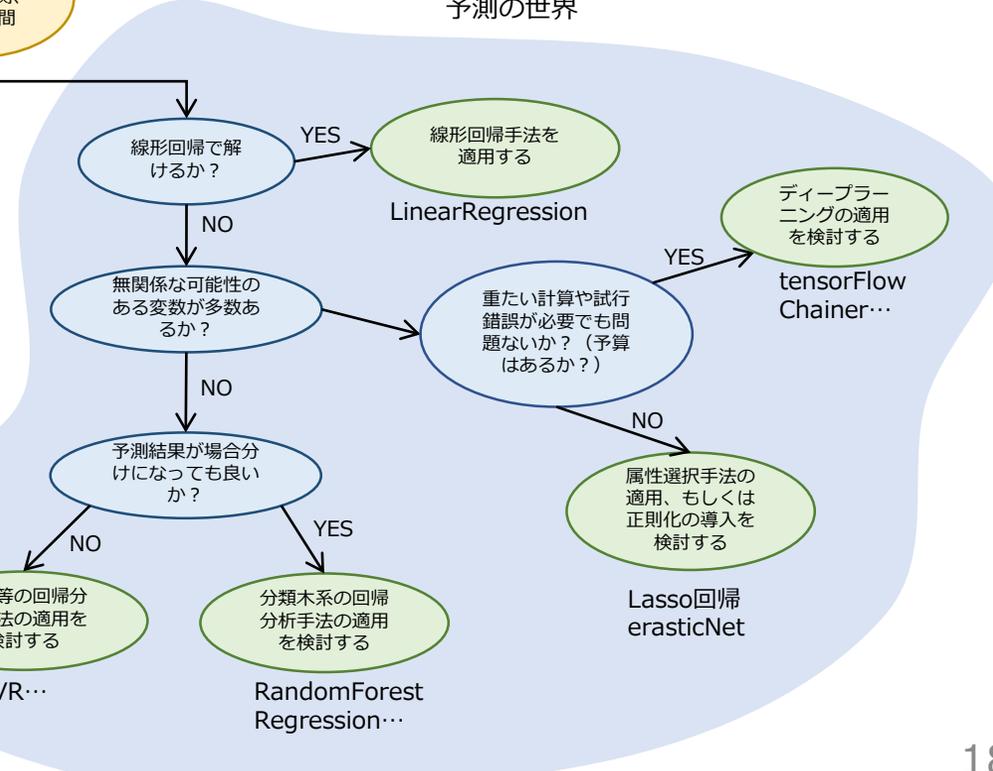
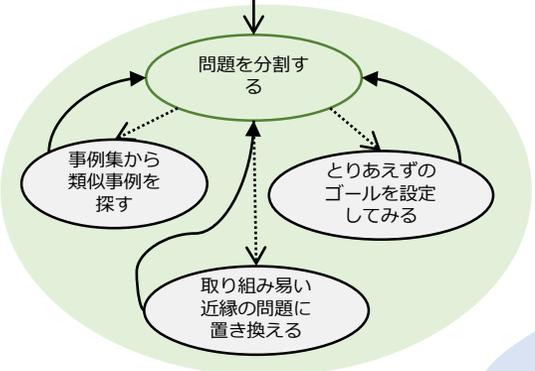
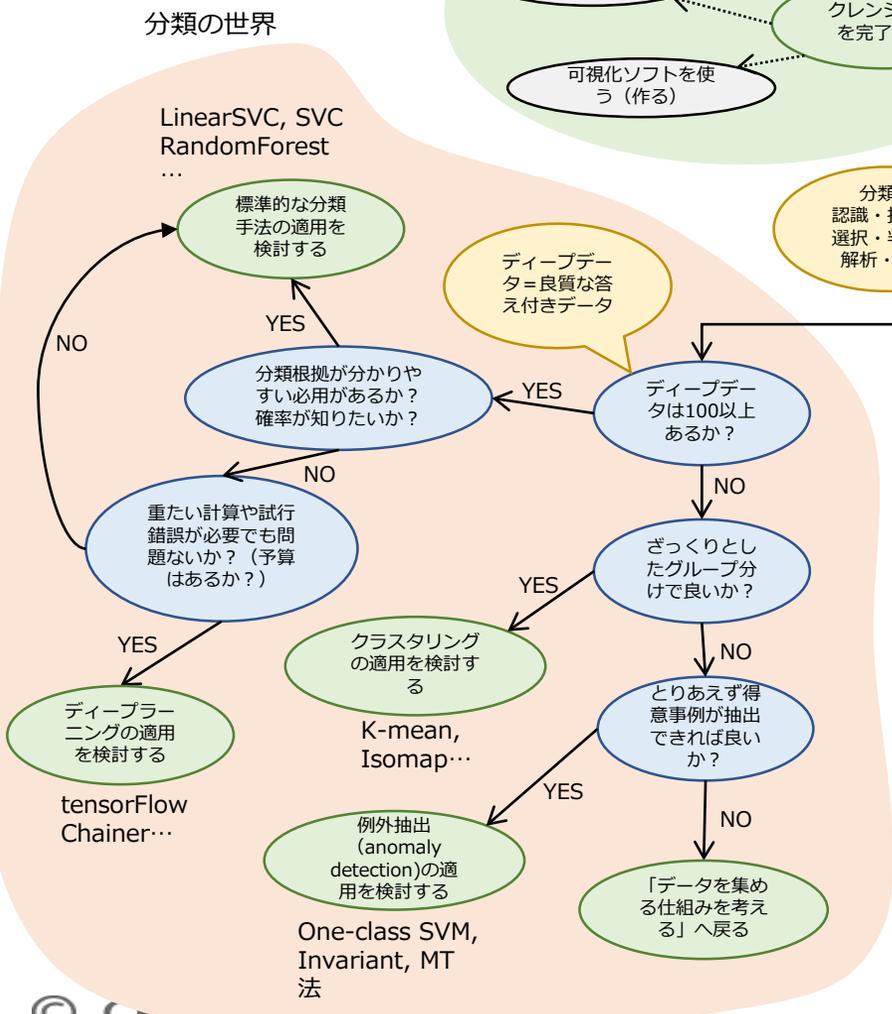
手元のデータ量や
やりたいことから
手法を選択できるフロー



有用なデータ収集方法を現場と一緒に考え・構築する段階

参考: scikit-learnのチートシート

scikit-learnチートシート
https://scikit-learn.org/stable/tutorial/machine_learning_map/index.html



いまのMLは何ができないか？

並 過去の類似事例を探したい。
気付き（インサイト）が欲しい。
例：操作者に設備に関する気付きを与える。

取り組みやすいが、使いこなせる現場が必須

良 チェック業務を楽にしたい。
例：大量の類似書類のミスを探す。
判断基準を統一したい。
例：設備の錆のランク判定を統一したい。
見落としを無くしたい。
例：大量の巡視ビデオのチェックをもれなく。

いまのMLが得意な分野

人間にもできることを代行させる

難 次の点検までの期間を見積りしたい。
めったに壊れないモノの寿命を知りたい。
会話の要約、申請書の自動作成。

かなり難度の高い応用
失敗覚悟で取り組むべし

要 データが無いので溜めたい。
答え付きデータ（ディープデータ）の蓄積を開始したい。
分類問題なのか、予測問題なのかわからない。

AI適用以前

苦手な問題の例

データのタイプ

- パターンに変換するのが難しいデータ
- 切れ目の分かりにくい時系列データ

判断基準・知識

- 判断のホワイトボックス化が必要な問題
(AIの判断理由を説明すべき問題)
- 常識と経験的知識の連携
- 言葉の意味・コンテキスト理解が必要な問題

答えの与え方

- 集団知を使っても答えが判然とは示せない問題
- 非明示的なシチュエーションで答えが変わる問題
- 人間の生物学的かつ非明示的な判断が影響する問題
(倫理など)

比較的近い将来にできそうなこと（5～10年先？）

- (1) **前処理がいらなくなる。**
（自動クレンジング、特徴量の自動抽出）
- (2) 教師データが少なくても、**とりあえずの答え**が出るようになる。
（教師なし学習、アドホック学習）
- (3) 一度教えたら、**似たようなことはできる**ようになる。
（転移学習）
- (4) 仕上がりの良し悪しだけ言えば、**やり方は自分で学ぶ**ようになる。
（強化学習）
- (5) 複雑な問題も、目標を示せば、人間が教えなくても**勝手に成長**して良く解けるようになる。
（生成・敵対的ネットワーク）
- (6) 時系列やシナリオなど**連続的な関係性を上手に扱**えるようになる。
（RNN、LSTM）

なんちゃって

準汎用ML

RI 電力中央研究所