

# 機械学習 学習する人工知能

鹿島久嗣

京都大学 情報学研究科 知能情報学専攻

# 集合知システム分野研究室のフォーカス： 機械学習によるデータ解析を柱に基礎から応用まで幅広く

1. 機械学習・データマイニング手法の開発
  - グラフ構造の機械学習、高次元小サンプルからの学習、...
2. データ解析の先進応用（ビジネス、サイエンス）
  - マーケティング、ヘルスケア、システム異常検知、...
3. 人とAIの協働問題解決：ヒューマン・コンピューテーション
  - クラウドソーシング、データ解析コンペティション、...

# 機械学習：

## 第3次人工知能ブームを支えるコア技術

---

- 近年の“人工知能”の成功：
  - クイズ王に勝利した質問応答システム
  - プロ棋士に勝利したコンピュータ将棋・囲碁
- 今回の“人工知能”ブームは機械学習に支えられている
  - とりわけ深層学習がブームを牽引



# 機械学習には何ができるか： 予測と発見

## ■ 機械学習の用途：予測と発見



### 予測型機械学習（≡教師付き学習）

- 過去～現在のデータをもとに、将来のデータについての予測をおこなう
- 「これから何が起こるのか？」



### 発見型機械学習（≡教師なし学習）

- 過去～現在のデータをもとに、何らかの知見を得る
- 「いま何が起きているのか？」

# ひろがる機械学習の応用： オンラインショッピングからシステム監視まで

## ■ マーケティング

- 商品推薦
- Web上の評判分析
- Web広告の最適化



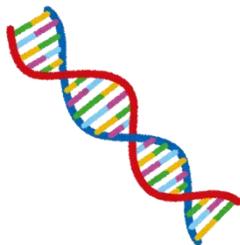
## ■ 金融

- 信用リスク評価
- 不正検出



## ■ サイエンス

- 遺伝子予測 (バイオ)
- 新規材料開発 (材料)



## ■ Web

- 検索
- スпам判定
- SNS



## ■ ヘルスケア

- 医療診断



## ■ マルチメディア

- 音声・画像認識

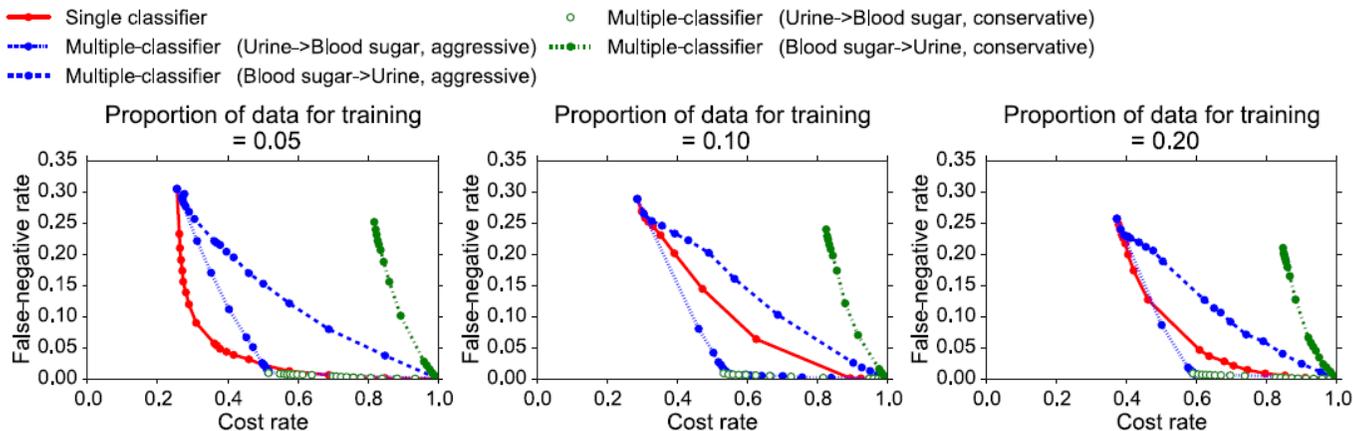
## ■ システム監視

- 故障・異常検知



# 予測型機械学習のヘルスケア応用： 途上国の医療支援

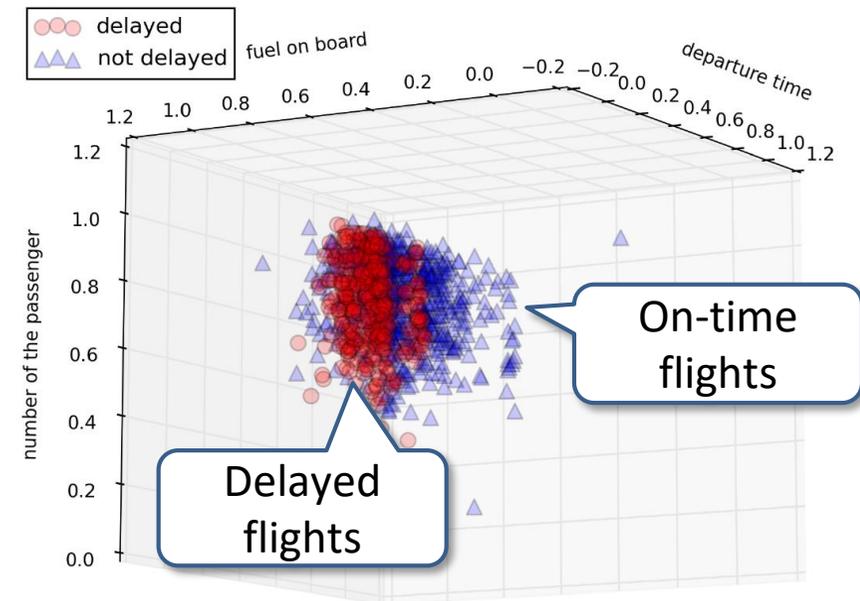
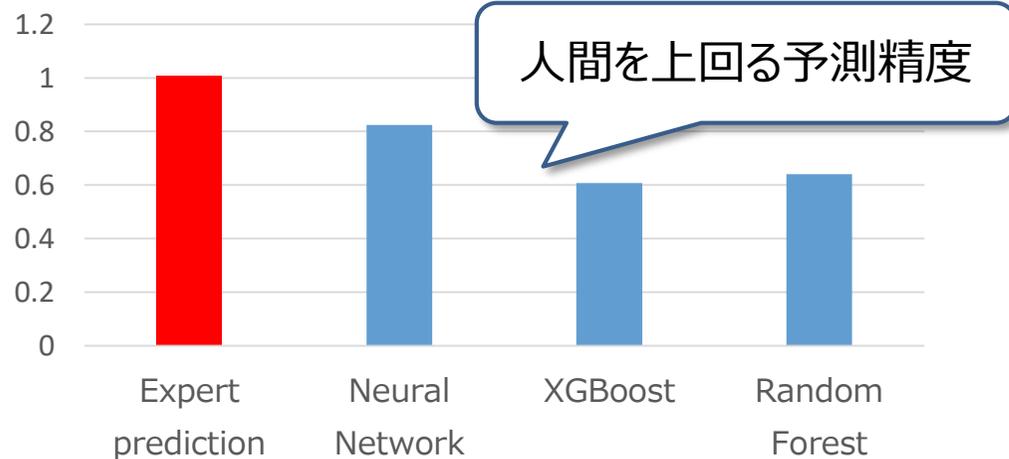
- 背景：途上国における生活習慣病の増加
- 医療支援：簡易的な検診と遠隔医療
- 機械学習による対象のふるい分け
  - 機械学習により、ふるい分けの精度を同等に保ちながら、50%の検診コスト削減
- 他、処方薬の推薦、予後の予測など



# 航空業界における予測型機械学習の応用： 予測モデル化による業務支援

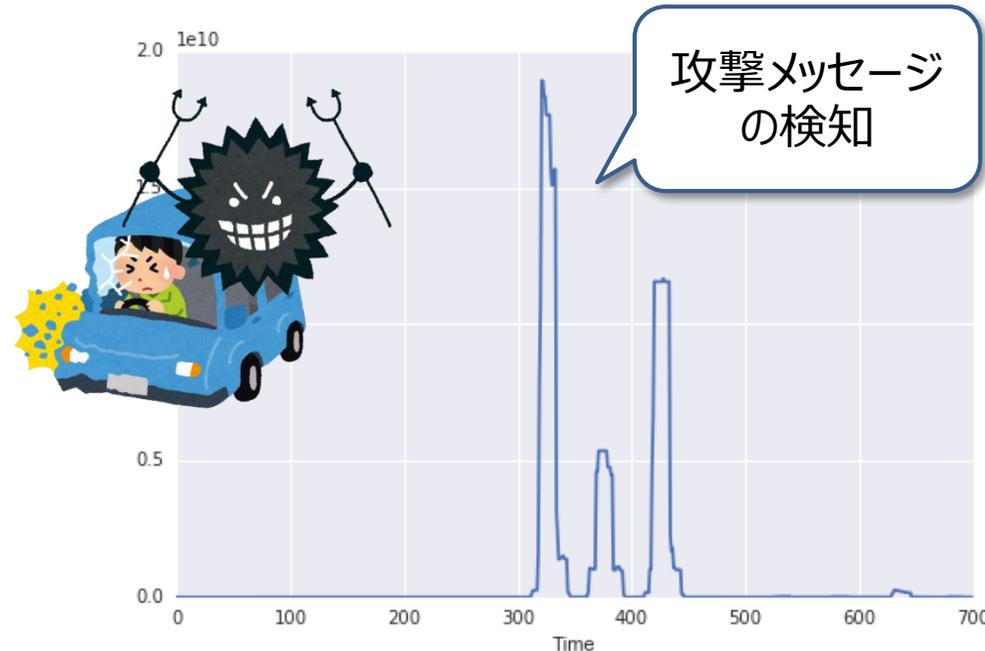
- 目的：機械学習によるコスト削減とサービス品質の向上
- 機械学習による燃料消費予測：  
専門家の予測を上回る予測精度
- 他、フライト遅延予測、機内販売需要予測、乗客数予測など

Prediction error (fuel consumption)



# 発見型機械学習の自動車セキュリティ応用： 車載ネットワークへの侵入検知

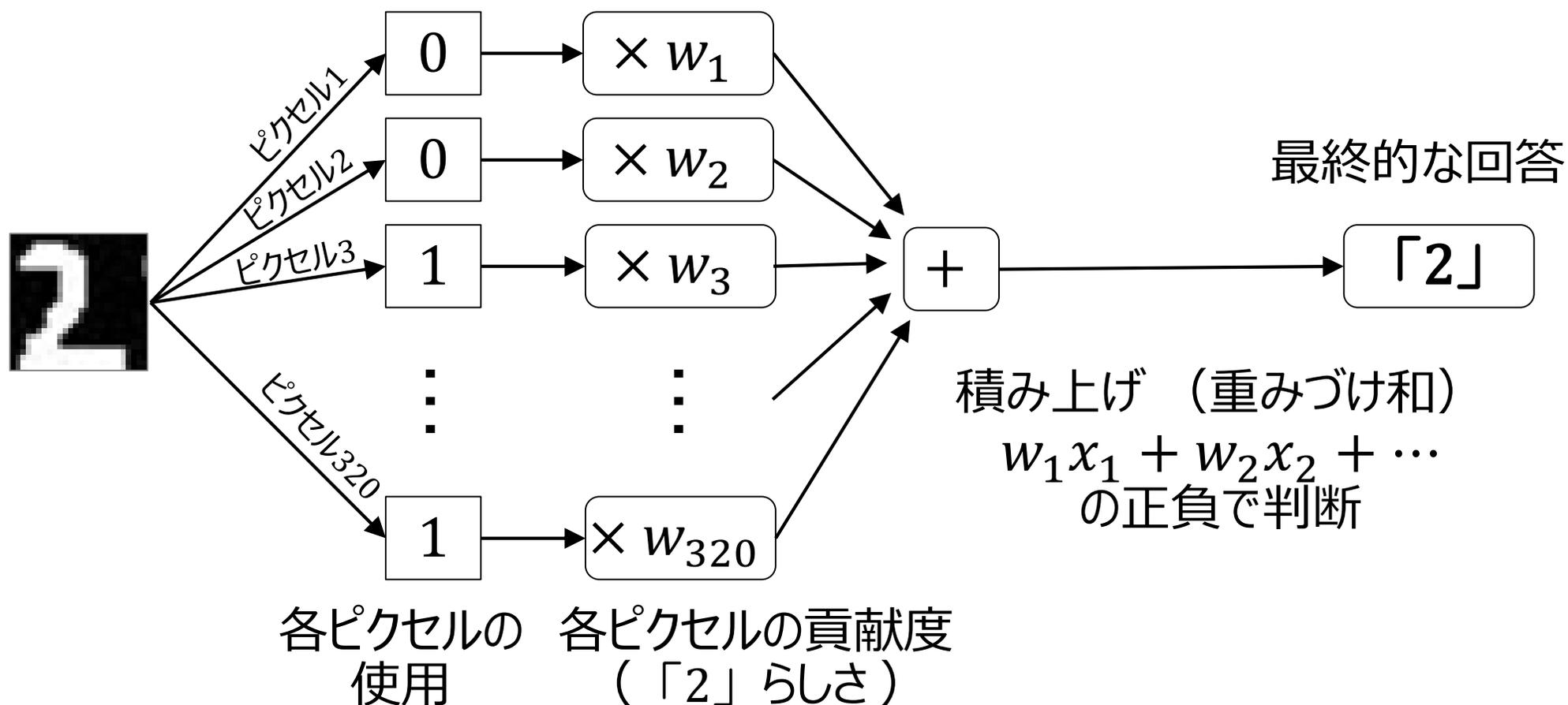
- 車載ネットワーク（CAN）に悪意あるメッセージを流すことで予期せぬ操作を起こす可能性
- 攻撃メッセージの自動検知
  - メッセージ頻度と内容に基づく統計的異常検知



# 典型的な予測モデル：

すべての証拠を重みづけて積み上げることで判断する

- 文字認識：各ピクセルの「2」らしさを重みづけ積上げる



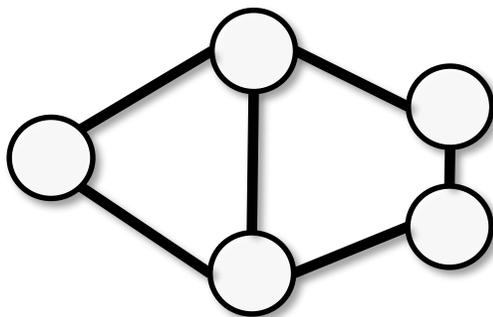
# データ解析のチャレンジ： グラフ構造をもったデータのモデル化

- 殆どの解析手法はベクトル形式（表形式）を仮定

ID	Name	Age	Sex	Address	...
0001	ABC	40	male	Tokyo	...
0002	XYZ	35	female	Osaka	...

ベクトル

- 非ベクトル形式のデータの扱いは自明でない
  - 中でもグラフはより一般的なデータ形式

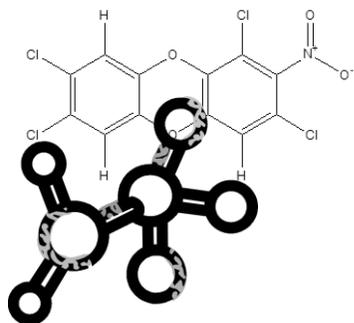


# グラフ構造データ：

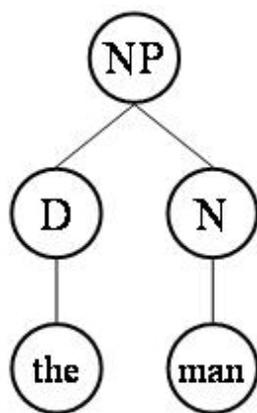
創薬, バイオ, Web, ... さまざまな場面で登場

## ■ さまざまなグラフ構造データ：

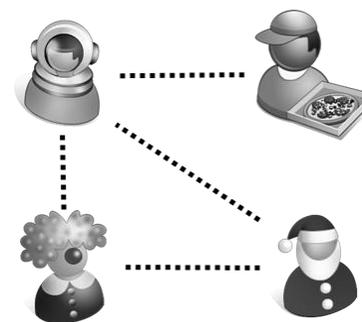
- ー 化合物、自然言語、ソーシャルネットワーク、HTML/XML、DNA、RNA、WWW、引用ネットワーク、生体ネットワーク、...



化合物



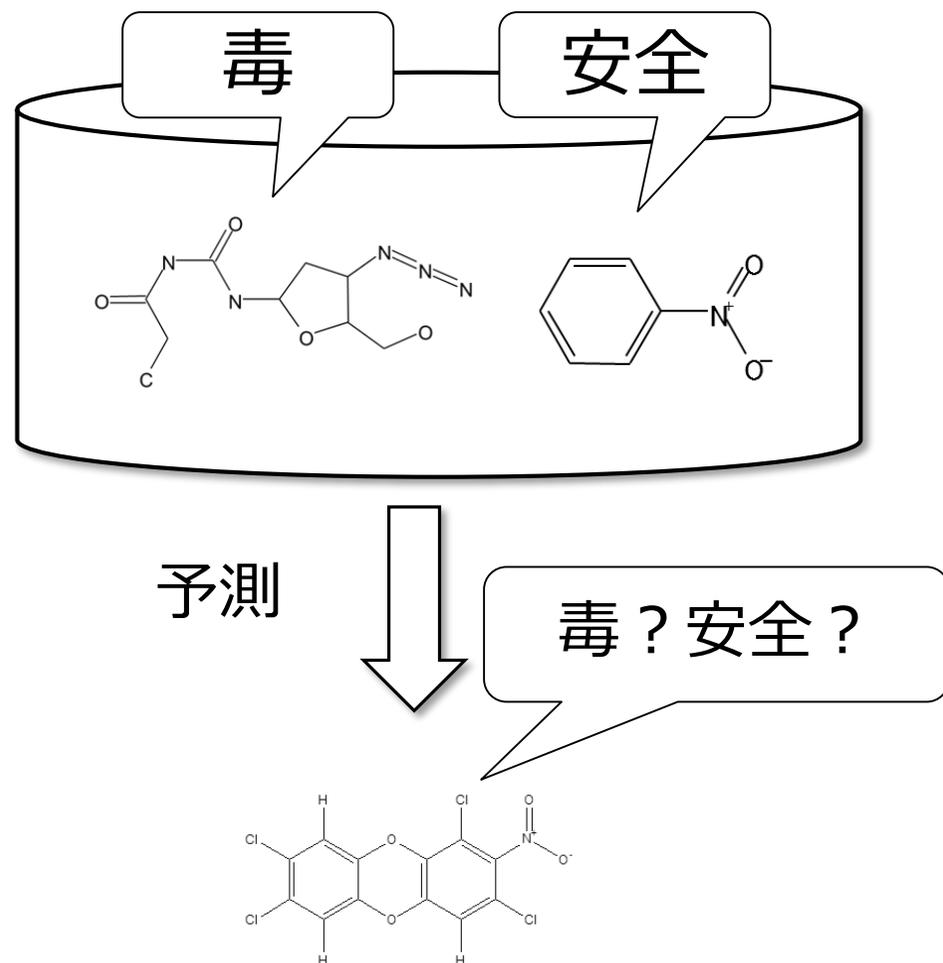
文（自然言語）



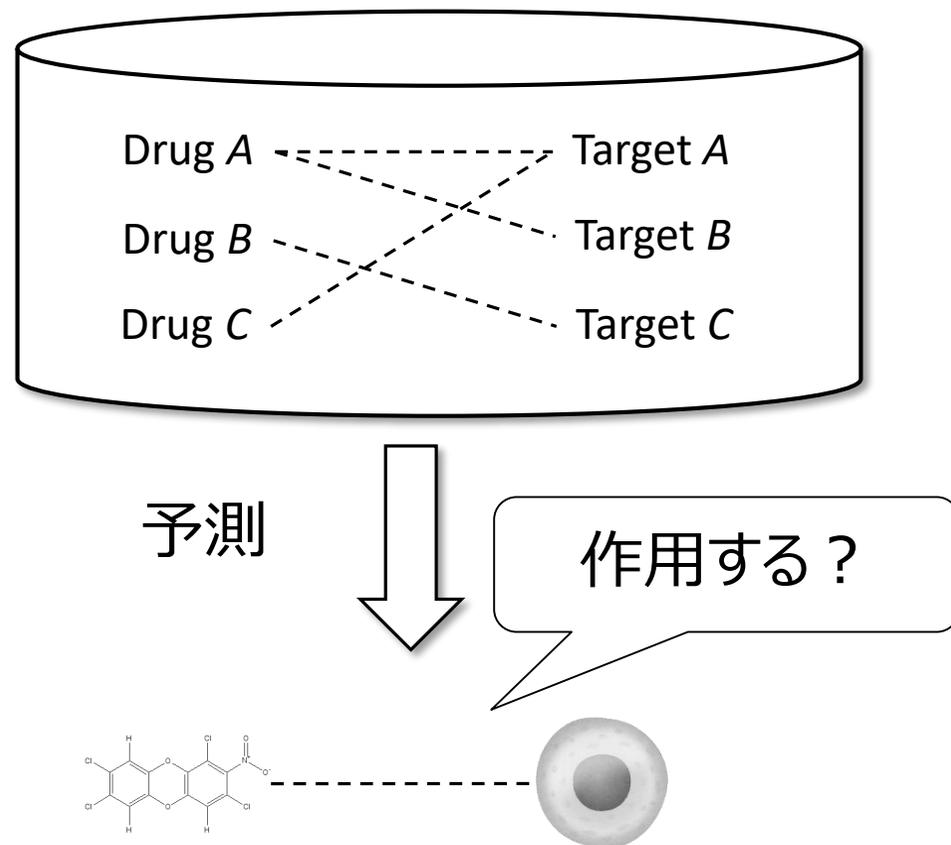
ソーシャルネットワーク

# グラフ構造機械学習： グラフ構造の特徴を捉えて予測

## 毒性予測



## 薬剤-標的 作用予測

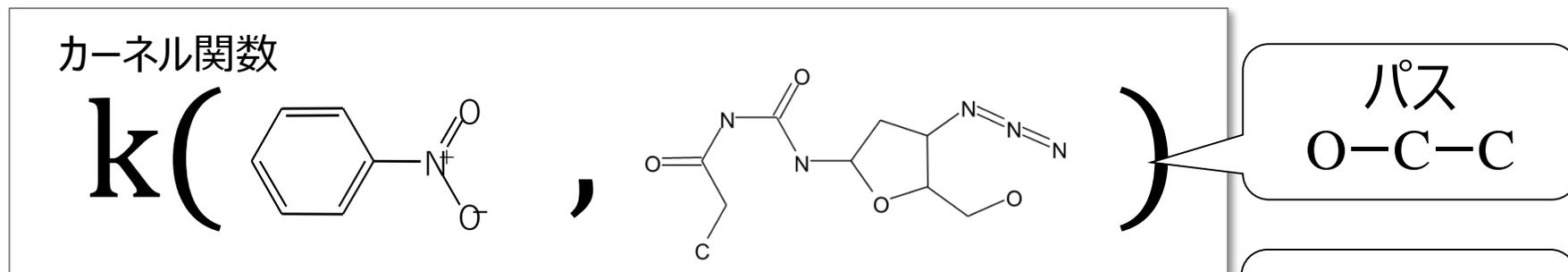




# グラフカーネル：

## 「初めての長時間（？）一般グラフ学習法」

- カーネル法: 類似度（カーネル関数）に基づく機械学習法
- グラフカーネル:  
2つのグラフ上に共通に含まれるパス数で構造的類似度を定義



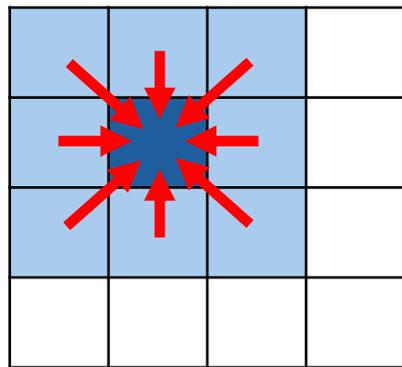
– 素朴な数え上げは不可能（ $\infty$ 個のパス）

– グラフ上でのランダムウォークによる暗黙的な数え上げ

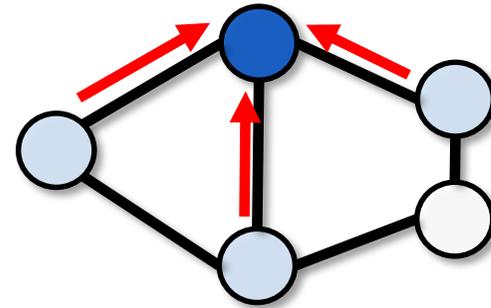
多項式時間

# 最近の展開： グラフ深層学習

- グラフ構造からの特徴抽出にニューラルネットワークを利用
- グラフ畳み込みニューラルネットワーク：
  - 画像畳み込みニューラルネットワーク：各ピクセルがその近傍ピクセルの情報を取り込む
  - グラフ畳み込み：各ノードが周辺ノードの情報を取り込む



画像畳み込み

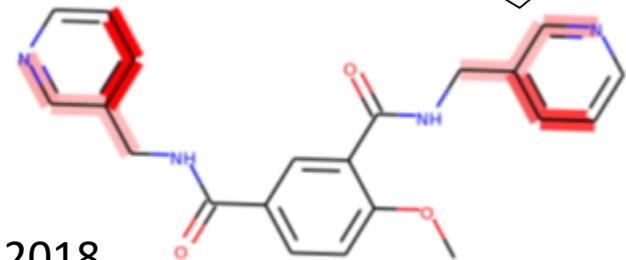


グラフ畳み込み

# 説明可能な機械学習： グラフに対する予測根拠の提示

- 機械学習による予測を盲信するリスク：
  - 機械学習（とくに深層学習）モデルはブラックボックス
  - 敵対的変動：データの僅かな変更で予測を変更可
- 自身の予測の根拠を与える機械学習モデルの必要性
  - グラフニューラルネットワークの予測の説明：  
判断に貢献する部分構造の抽出と可視化

この辺が  
効いてる！！



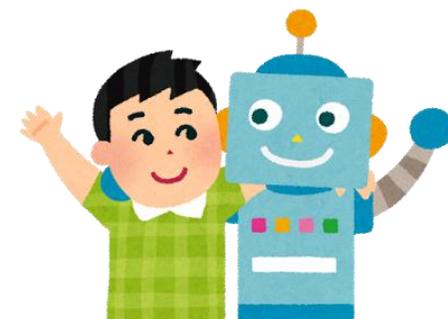
Akita et al., BayesGrad. ICONIP, 2018.

# 人工知能の限界： オープン・抽象的・文脈依存性の高い課題

- 人工知能技術の発展は目覚ましいが、魔法の道具ではない（「それ、E～AIで何とかありませんかねえ？」）
- 入出力が定義され、データが取得可能な領域においては機械が優勢
- 曖昧で文脈依存性の高い、抽象的な課題においては、人間の能力は柔軟性や創造性などにおいて上回る
- 人類と人工知能が協力して知的で困難な課題を解決するための方法論の開発が重要

# ヒューマンコンピューテーション： 人と機械の協働による新しい知能の実現

- 人間と機械のベストミックスは最強の人間と最強の機械を超える
  - チェスのフリースタイル（人間＋機械の自由なチーム編成）では人間＋機械のチームが最強
- ヒューマンコンピューテーションもしくは人間参加型（Human-in-the-loop）型AIと呼ばれるAIの新しい方向性の出現
  - コンピュータには困難な、さらには一人の人間では解くことのできない問題を、両者を適切に組みあわせて解決する



# クラウドソーシング： 不特定多数に仕事を依頼するしくみ

## ■ クラウドソーシングとは

「（インターネットを通じて）不特定多数の人に仕事を依頼すること、もしくはその仕組み」一般を指す言葉

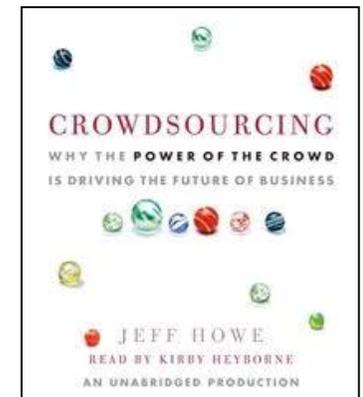
⇔ アウトソーシングの委託先は素性の知れた特定の相手

## ■ クラウドソーシングのメリット：

- 量：大量の「人間の知的作業」にアクセスできる
- 質：群衆の叡智の利用  
（「三人寄れば文殊の知恵」）



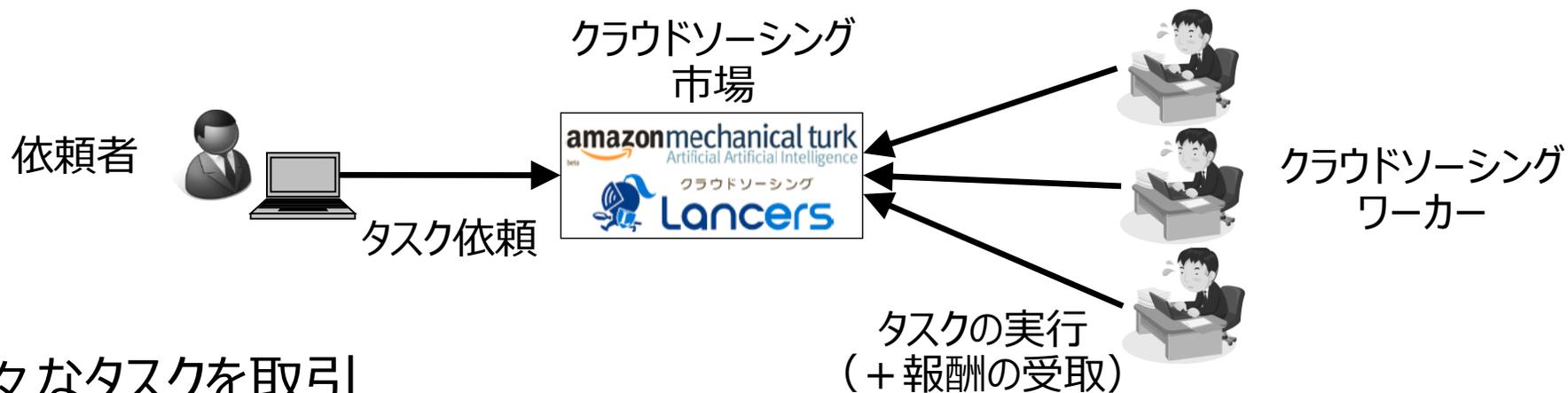
Howe. Crowdsourcing. 2004.



# クラウドソーシングプラットフォームの浸透： ヒューマンコンピューテーションのプラットフォーム

## ■ クラウドソーシング市場の出現

- 海外：Mechanical Turk, oDesk, ...
- 国内：ランサーズ, クラウドワークス, ..., ...



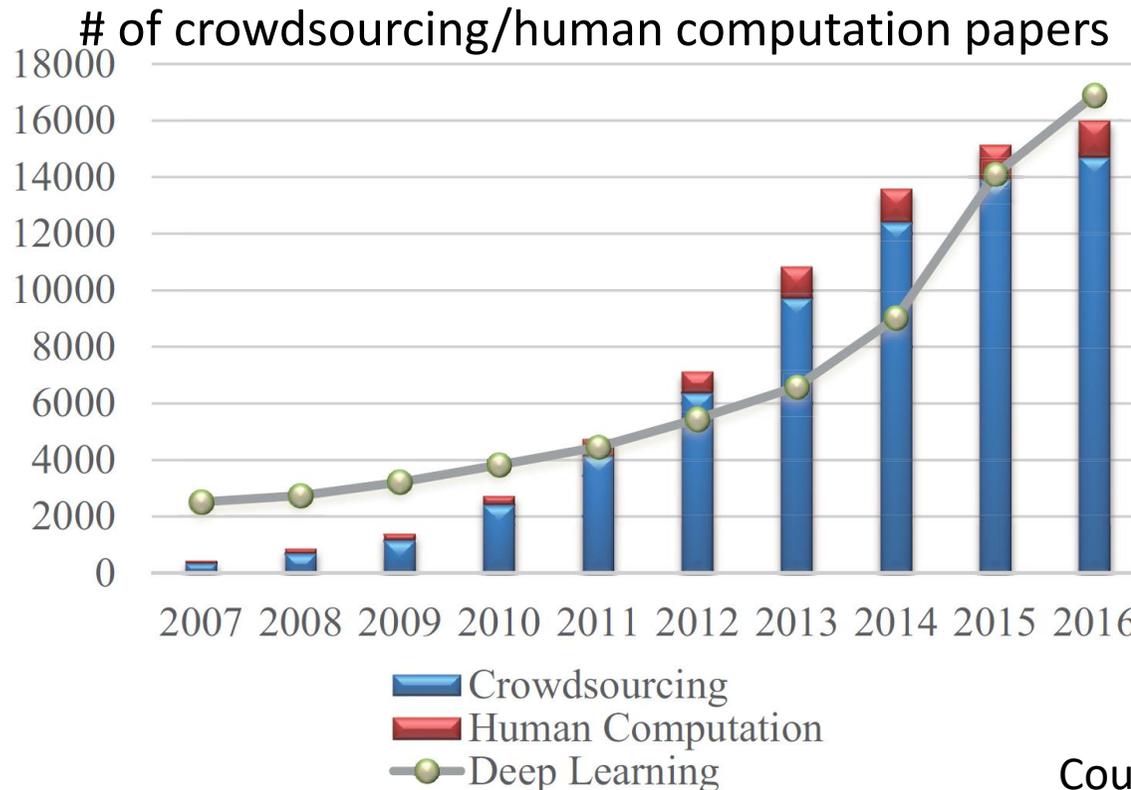
## – 様々なタスクを取引

- マイクロタスク：画像識別, 情報抽出, データ確認, 音声書き起こし, ...
- 複雑なタスク：デザイン, システム開発, ...

## ■ ヒューマンコンピューテーションのプラットフォーム

# 研究のトレンド： クラウドソーシング・ヒューマンコンピューテーション研究の増加

- 2005: Amazon Mechanical Turk の登場
- 2006: 「クラウドソーシング」「ヒューマンコンピューテーション」の提唱
- 2013: 1st conference on human computation and crowdsourcing (HCOMP)



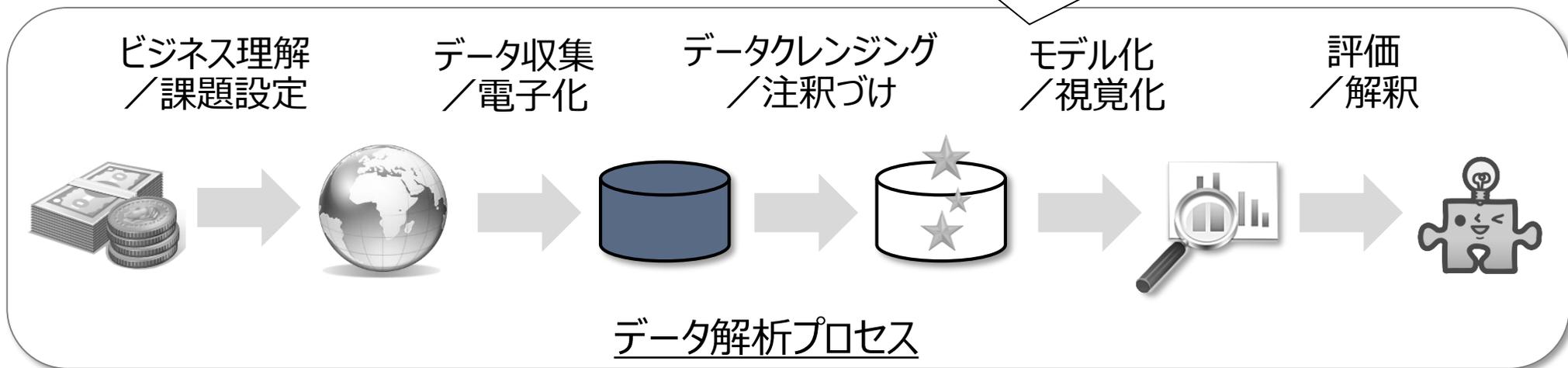
Counted using Google Scholar

# データ解析の労働集約性：

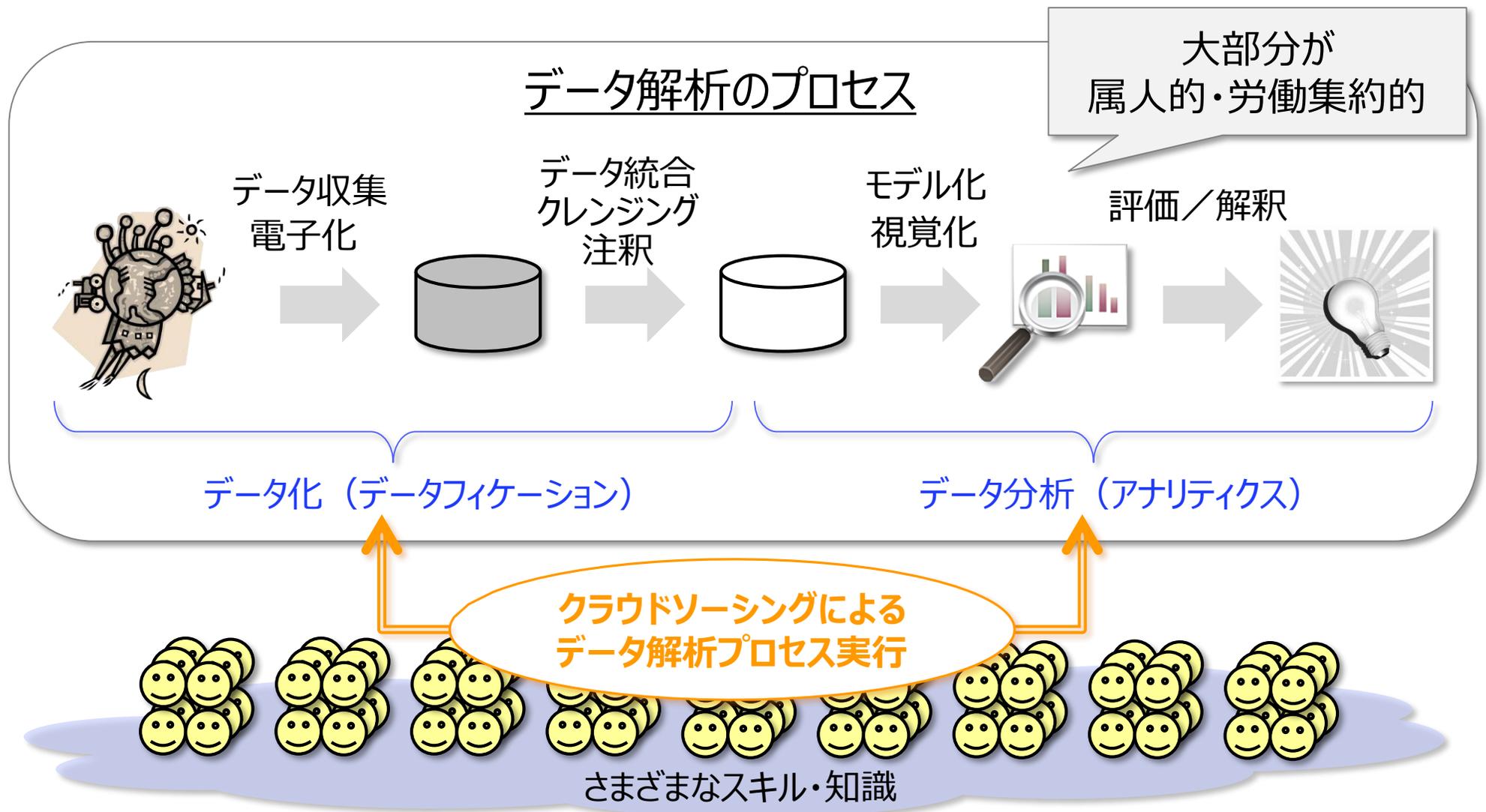
データ解析プロセスの多くの部分は依然として人力

- 機械学習等によるデータの自動解析は、しばしばデータ解析の中心として捉えられる
- データの収集や洗浄／結果の解釈などを含むデータ解析のプロセス全体を見渡すと、多くの部分が人間に依存する
- データサイエンティストの不足

自動化できるのは全体の2割程度



# クラウドソーシングによるデータ解析： 知識・スキル・労働力を集め解析プロセスを実行する

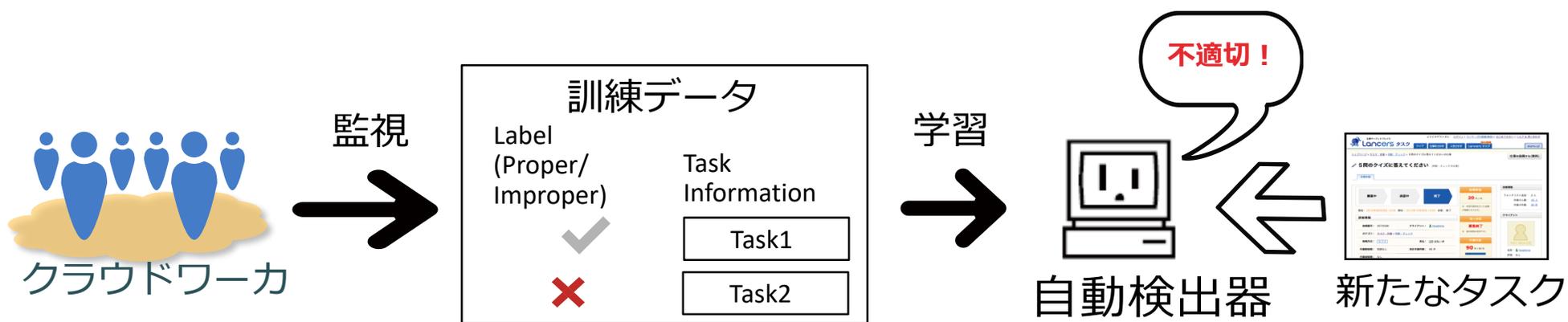


# 人間参加型（Human-in-the-loop）機械学習： クラウドソーシングによる機械学習用データの収集

- 機械学習にはデータが必要
  - AIの勝利条件：
    - 入出力の形式的な定義
    - データの取得可能性
- 人間がデータ収集を助ける
- クラウドソーシングを利用したデータ収集：
  - 正解となるラベル（学習データ）の収集
  - 適切な特徴量の設計・収集

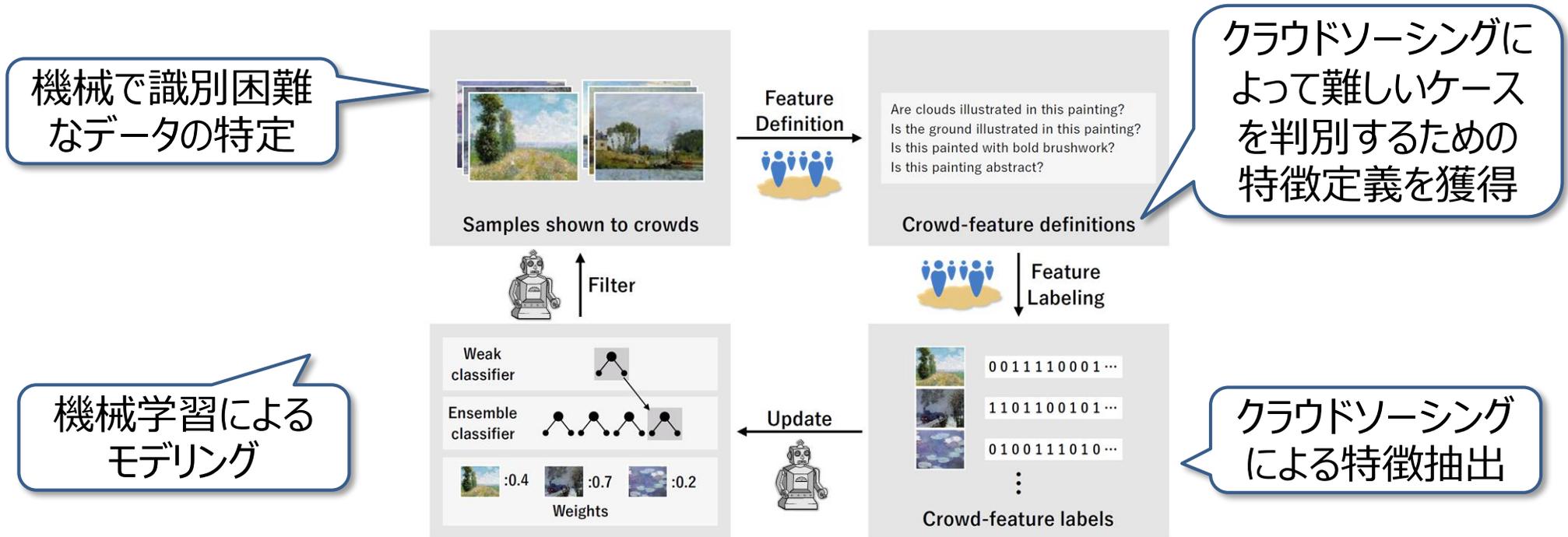
# クラウドソーシングによる学習データ作成： クラウドソーシングプラットフォーム運営サポートの例

- プラットフォームの運営のための課題：健全性の確保
  - 不適切タスク：個人情報収集, 特定のサービスへの登録依頼, SNSへの投稿, ...
- “自警団”：クラウドワーカーに不適切タスクの検出を依頼
  - ワーカーの判断を訓練データとして自動判別器を学習



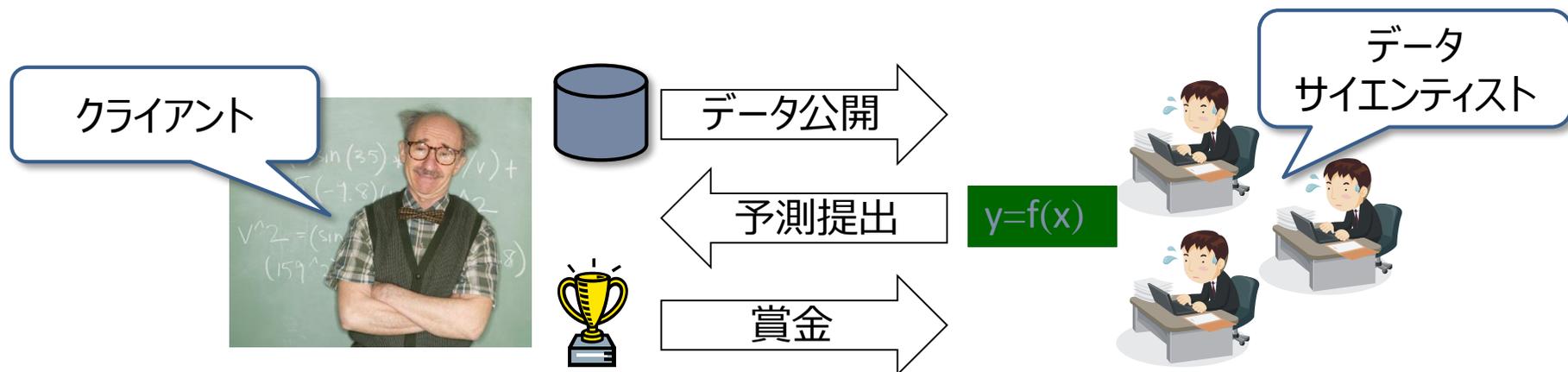
# クラウドソーシングによる特徴生成： 機械学習のための特徴を人間が生成

- 機械学習のアルゴリズムの中に人間が参加することで、機械で困難な特徴の発見・抽出を行う
- ブースティングの枠組みを利用して逐次的に特徴生成



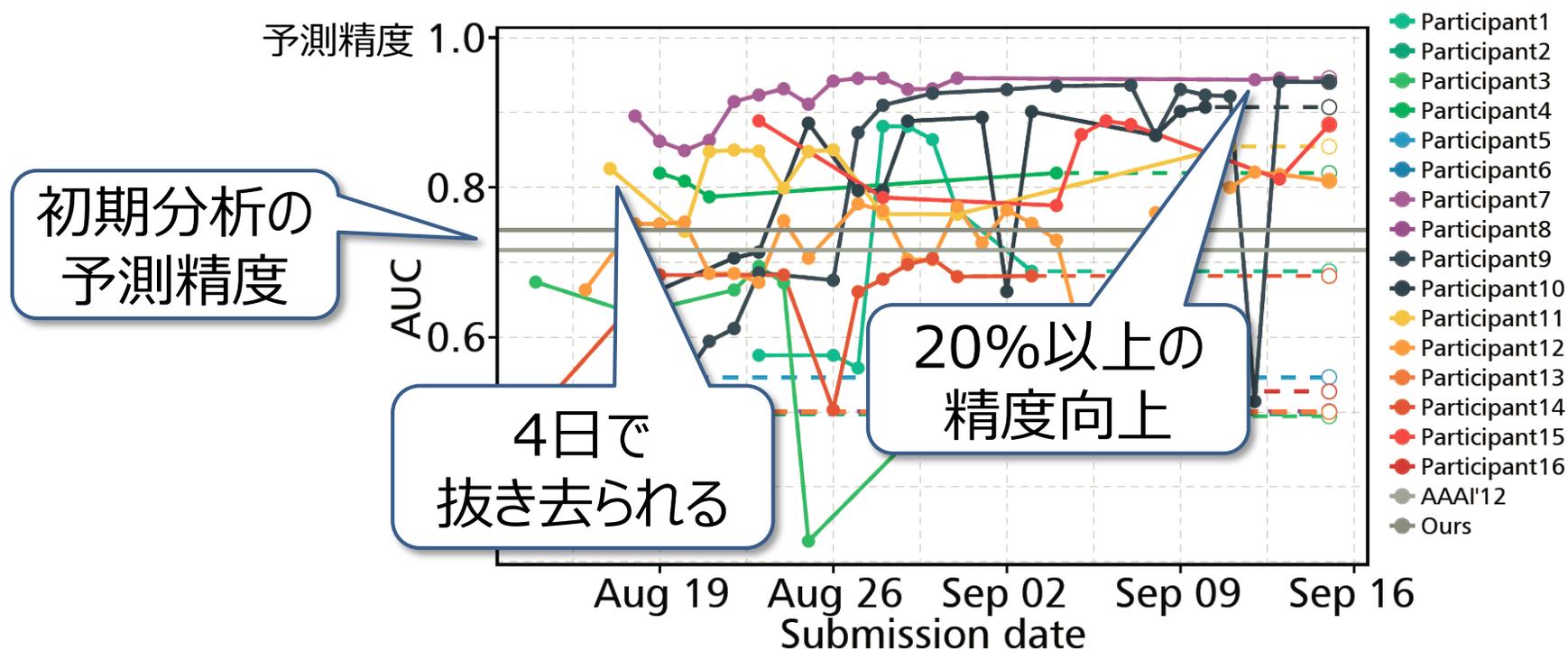
# データ分析のクラウドソーシング： データ解析コンペティション

- 自動化の進むモデリング部分でも労働集約性は高い
  - データに合ったモデルを（人手で）広範囲に探索する必要
- データ解析コンペティション：モデリングのクラウドソーシング
  - データを公開し、結果（予測精度）を競う
- プラットフォームの出現：[kaggle](#)  [DeepAnalytics](#)  [UNIVERSITY OF BIG DATA](#)



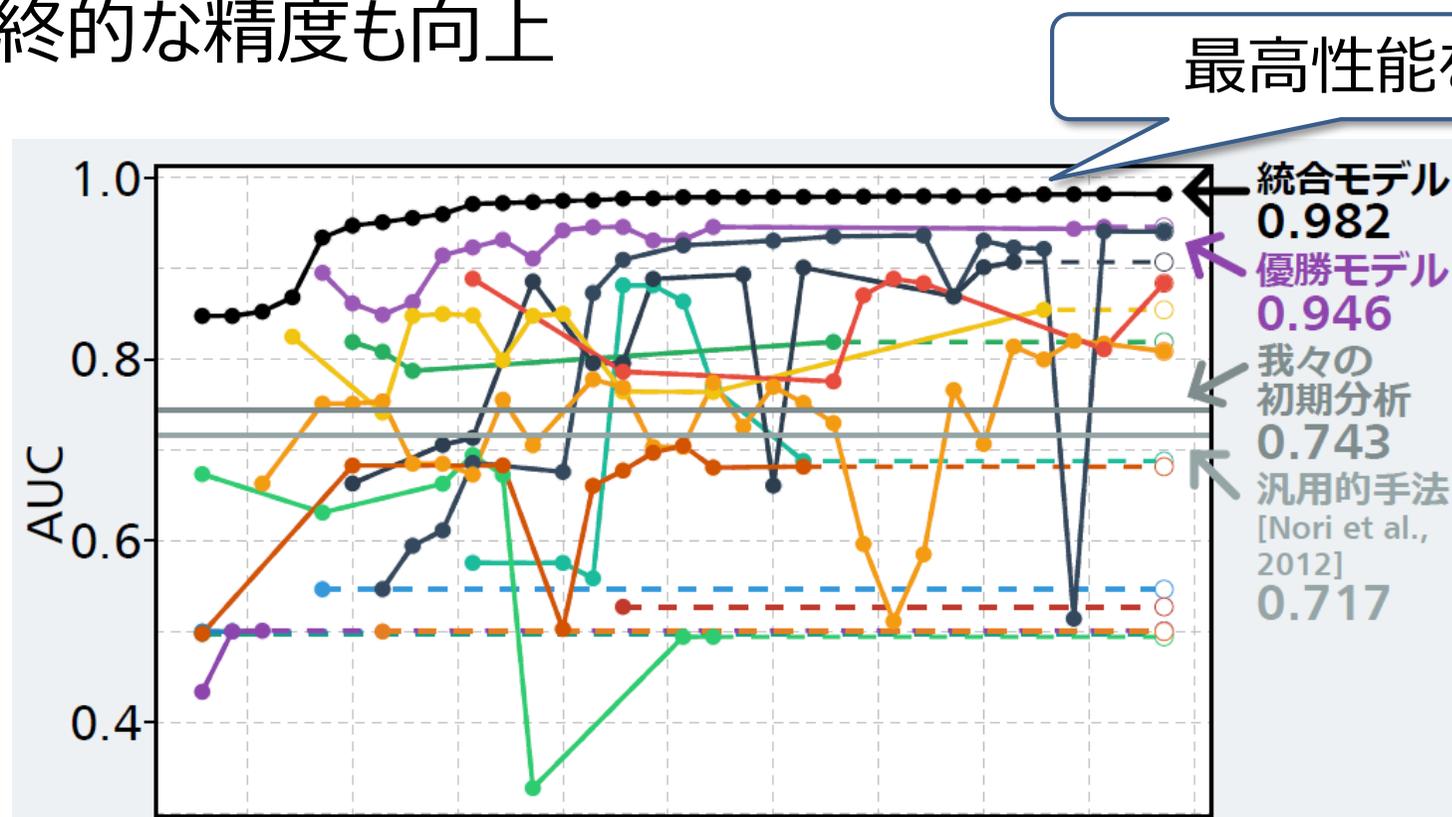
# 予測コンペティションの結果： 短期間で“プロ”を超える精度を実現

- Wikipediaのリンク予測を題材にコンペ開催（16名が参加）
- 短期間で“プロ”を遥かに超える予測精度を達成
  - 初期分析結果を4日目で抜き、最終的に20%以上の精度向上



# 機械学習と集合知の融合： 予測器の組み合わせによって最高精度を達成

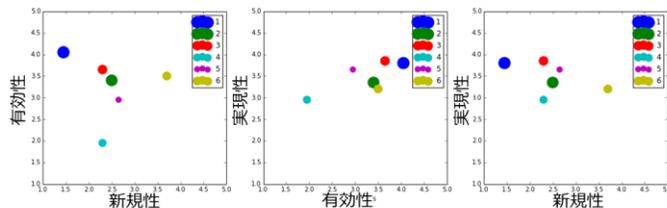
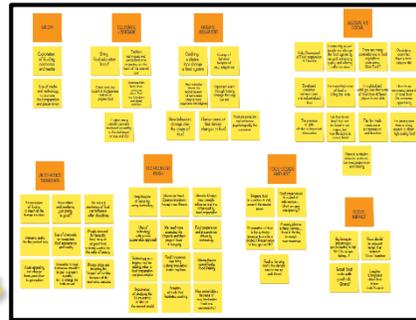
- 提出された予測器を機械学習によって組みあわせる
  - 5日目で、個人の最高精度と同程度の予測を達成
  - 最終的な精度も向上



# 人とAIによる一般問題解決： 不特定多数の力で問題解決のアイデアを探る

- アイディアの生成：クラウドソーシングでアイデア収集
- アイディア組織化：アイデア集合の整理・優先度付け

Problem finding, ideation,  
and organization with  
crowds



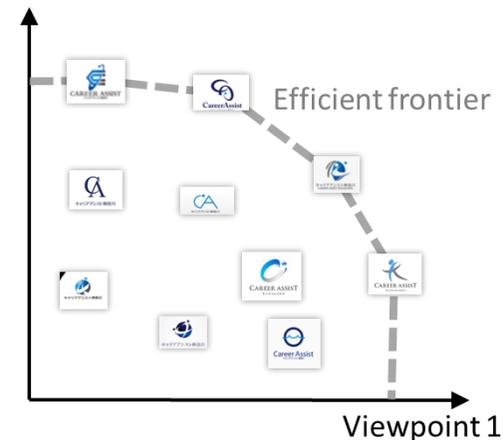
番号	グループのラベル	番号	グループのラベル
1	カンニング防止のための徹底した監視	4	生徒の意識改善策
2	物理的なカンニング防止策	5	席配置改善グループ
3	持ち物制限・教室変更グループ	6	試験問題への工夫

問題を2種類作成し、教室の机の列ごとに異なる問題を配布する  
問題は変えずに大問の配列を数パターン作り、席の列ごとに違う配列の解答用紙を配る  
問題のストックを出来るだけたくさん用意して、各自のテスト内容を変える

「試験問題への工夫」

カンニングをした生徒は即退学というルールを作る  
普段から生徒に口やかましく、カンニングは、犯罪だと伝えておく

「生徒の意識改善策」

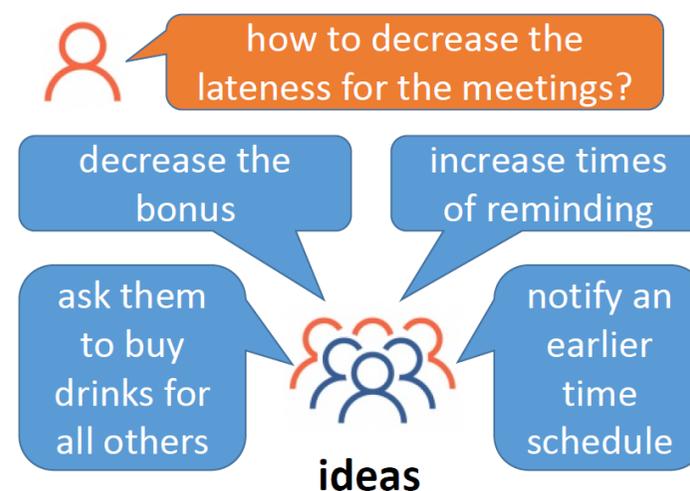


Viewpoint 1

# アイデア組織化：

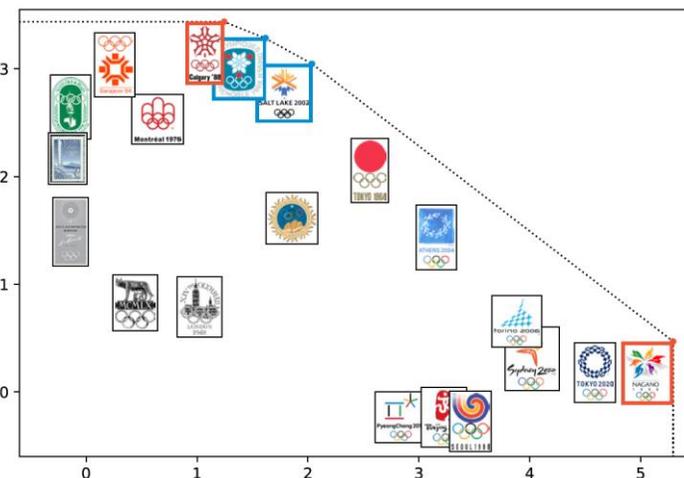
## クラウドソーシングによるアイデアの整理と順序付け

- 集めたアイデアやデザインを元に意思決定を行いたい
- クラスタリング：アイデアの集まりを理解する
- ランキング：アイデアの優先度付け
- 完全自動化は困難 → クラウドソーシングによる判断



# 現在の研究フォーカス： 人々をより賢い・正しい方向に導くAI

- AI困難（AI-hard）な問題への一般的アプローチ
- 多数の人間によるオンライン問題解決手法
  - 様々な観点からのアイデア組織化
    - 異なる観点をもつ参加者の評価をもとに優先度マップを作成
- AIが人を賢くする
  - ラーニングアナリティクス



# まとめ： 機械学習と集合知

---

- 機械学習：人工知能ブームを支える技術
- グラフ機械学習：複雑な構造データへの挑戦
- AI困難問題：オープン・抽象的・文脈依存性の高い課題
- ヒューマンコンピューテーション：人工知能と人間知能の共同協調問題解決
- クラウドソーシング：集合知によるデータ解析と問題解決