

一般教養としての AI と その周辺

九州大学 システム情報科学研究所
准教授 稲永 俊介



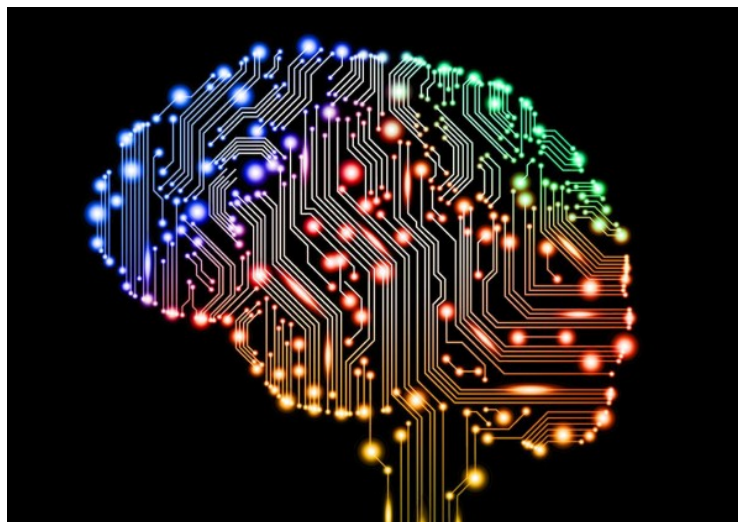
人工知能とは？

人工知能??

- 近年、人工知能 (AI) という言葉がニュース等で盛んに使われるようになっている。
 - 「A 社が〇〇ができる人工知能を開発した」
 - 「B 社が人工知能を搭載した××を発売する」
- 一部では、将来的に、人工知能が人間の職業を奪う可能性がある」と報じられている。
- では、そもそも人工知能とは一体何なのか？

人工知能とは

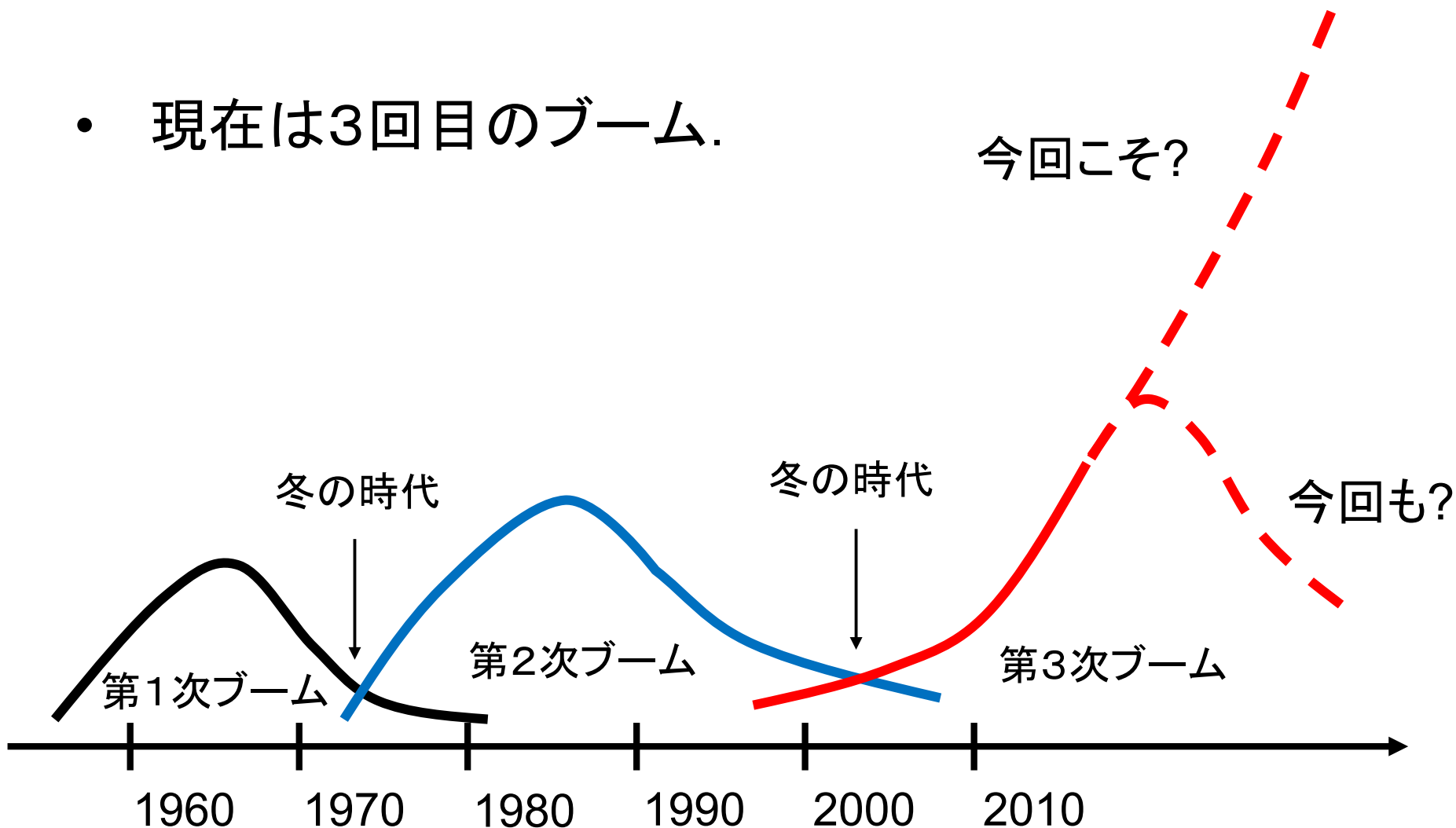
- 人工知能（Artificial Intelligence, AI）とは、コンピュータ上で人間と同様の知能を人工的に実現させようという**試み**、或いはそのための一連の基礎技術のこと。（Wikipedia 参照）



画像: <https://www.technologyreview.com/s/601519/how-to-create-a-malevolent-artificial-intelligence/>

人工知能の歴史

- 現在は3回目のブーム.



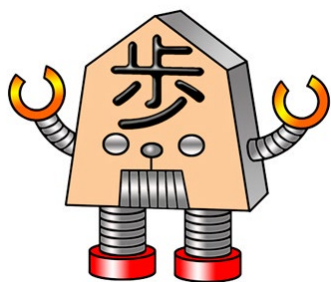
図は「人工知能は人間を超えるか」松尾豊(著)を参考に作成

最近の動向

人工知能に関する最近のニュース 1

【将棋】

- 2010年10月：清水市代 女流王位・女流王将とあから2010 が対戦し，あから2010 が勝利.
- 2012年1月：米長邦雄 永世棋聖とボンクラーズが対戦し，ボンクラーズが勝利.
- 2014年3～4月：第3回将棋電王戦において，コンピュータ将棋がプロ棋士に4勝1敗で勝ち越し.



あから2010



ボードゲームの複雑さ

- 探索空間(第一手から最終局面までの手順の種類数)が大きいほど, そのゲームは難しい.
- 探索空間のサイズ

オセロ	$\sim 10^{60}$
チェス	$\sim 10^{120}$
将棋	$\sim 10^{220}$
囲碁	$\sim 10^{360}$

- 探索空間のサイズで比較すると,
囲碁は将棋の 10^{140} 倍(1兆の10乗以上)複雑!

人工知能に関する最近のニュース 2

【囲碁】

- 2015年までは、最も強い囲碁プログラムでも、人間のアマチュアレベルに達するのがやっと。
- アルファ碁 (Google DeepMind 社) の登場
 - 2015年10月: プロ棋士に勝利 (5-0)
 - 2016年 3月: 元世界王者 イ・セドルに勝利 (4-1)

衝撃!



人工知能に関する最近のニュース 3

【自動運転】

- 2015年10月, 米テスラ社が同社の車両に搭載可能な半自動運転機能（オプション）を発売.
- 2016年10月, 同社は完全自動運転機能（オプション）を発売.



画像 <http://www.caranddriver.com/tesla/model-s>

人工知能に関する最近のニュース 4

【クイズ番組】

- 2011年, IBM が開発した質問応答システム「ワトソン」がアメリカのクイズ番組 Jeopardy! で人間のクイズチャンピオンと対戦して勝利.



画像 <https://i.ytimg.com/vi/P18EdAKuC1U/maxresdefault.jpg>

人工知能に関する最近のニュース 5




【機械翻訳】

- 機械翻訳(自動翻訳)は人類長年のテーマ.
 - 外国語が苦手な人が多い日本人は特に…
- 類似言語間
(英語⇔フランス語, 英語⇔ドイツ語 など)
の翻訳は数年前からかなりの精度を誇っている.
- ごく最近になって, 英語⇔日本語などの
非類似言語間の翻訳精度も向上した.

人工知能に関する最近のニュース 5

【機械翻訳】

Google 翻訳 (インターネット上で, 誰でも無料で利用可能)

英語 ▾	  	日本語 ▾	
Kyushu University is a Japanese national university, located in Fukuoka, Kyushu.		九州大学は、九州の福岡にある日本の国立大学です。 Kyūshūdaigaku wa, Kyūshū no Fukuoka ni aru Nihon no kunitachidaigakudesu.	

人工知能の現状

- 現在の人工知能技術は、以下の2つの要素技術を組み合わせて作られている。

➤ 機械学習



画像 http://www.cs.toronto.edu/~urtasun/courses/CSC411_Fall16/CSC411_Fall16.html

➤ ビッグデータ解析

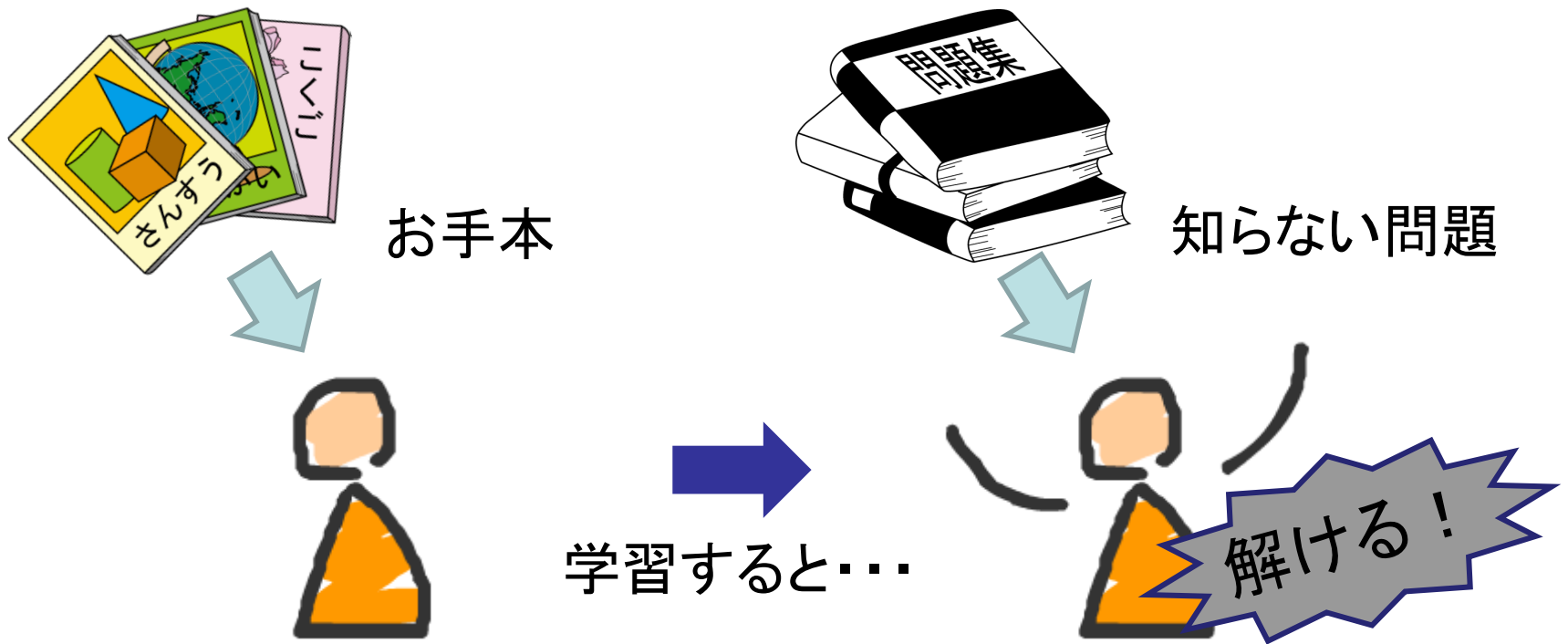


画像 <http://www.trickymindsolution.com/big-data-analysis-services.html>

機械学習

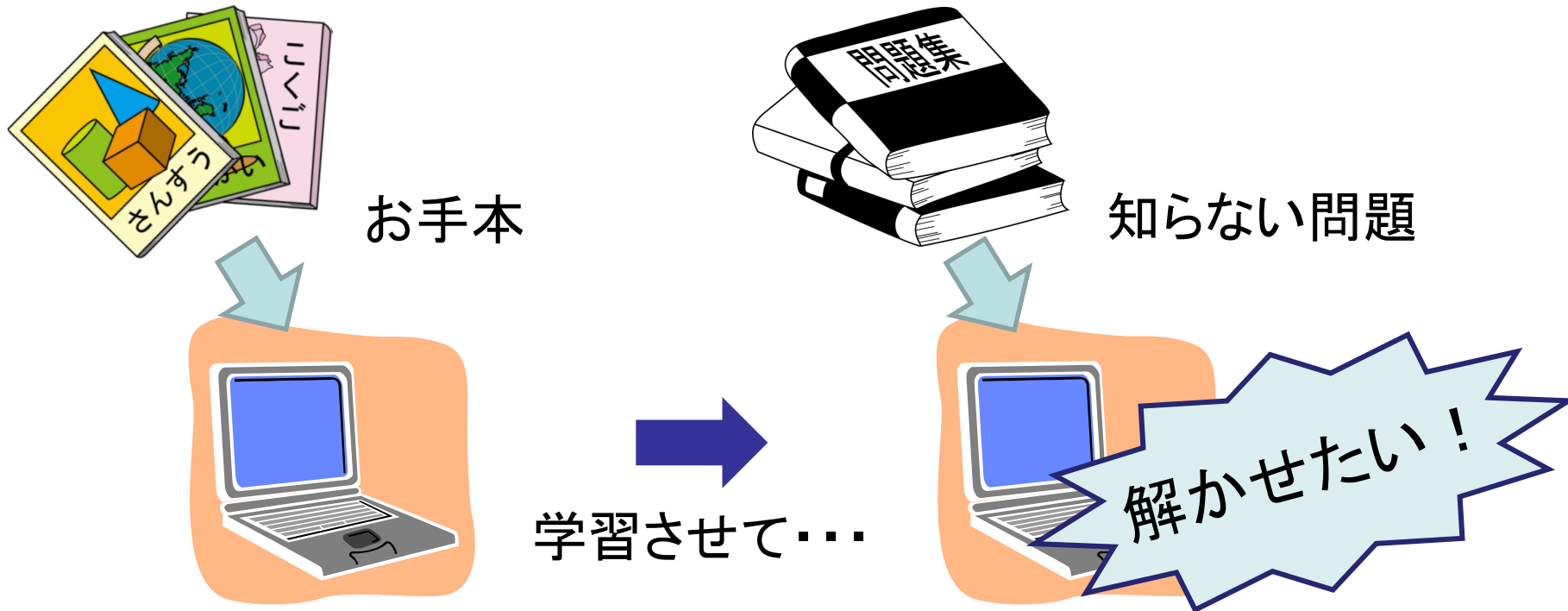
学習

- 学習：学んで知識を理解したり記憶したりすること (Wikipedia)



機械学習

- 機械学習: 人間が自然に行っている学習能力と同等の機能をコンピュータで実現させるための技術 (Wikipedia)



機械学習の基本的な考え方

機械学習の基本的な考え方を

- 分類問題
- オンライン予測問題

を題材にして簡単に解説する.

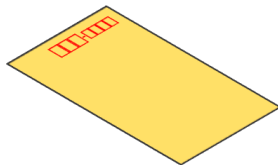


分類問題

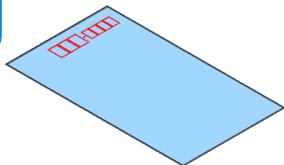
- 手法: コンピュータは, まず, 既知のデータ (訓練データ) を用いて, データを分類するためのルールを学習する.
- ゴール: 未知のデータが与えられたとき, それを正しく分類したい.
- 以降, 分類問題のわかりやすい例として, スпам (迷惑メール) の分類を考える.

スパム分類問題

スパム

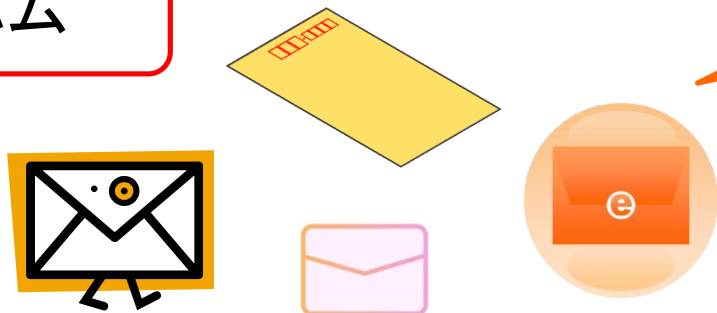


スパムでない



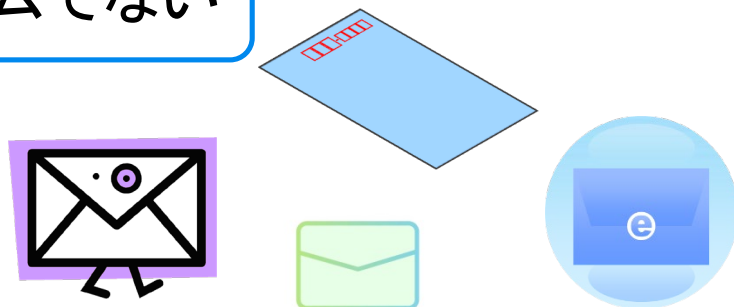
スパム分類問題

スパム



- 利用料が...
- 高収入が...
- 友達申請が...
- 出合いが...

スパムでない



こんなのが
スパム！

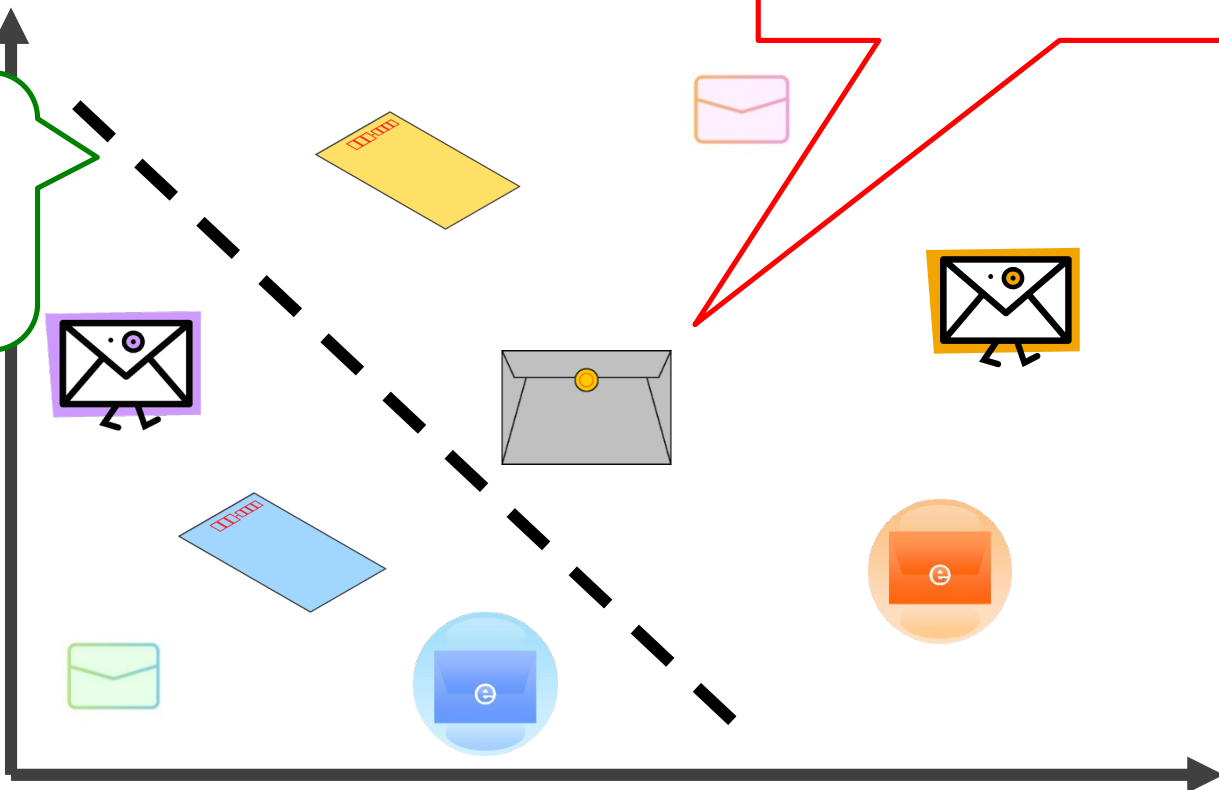
スパム分類問題

- スпамと非スパムを分類するために、分類のルールを定める必要がある。
- スпамに特徴的だと思われる語句の例
 - 利用料
 - 高収入
 - 友達申請
 - 出会い
- そこで、(いくつかの)単語の出現回数でスパムと非スパムを分類することを考えてみる。

スパム分類問題

単語Aの数

スパムと
非スパムの
分類線



簡単なサイドビジネスで
高収入を得ませんか?...

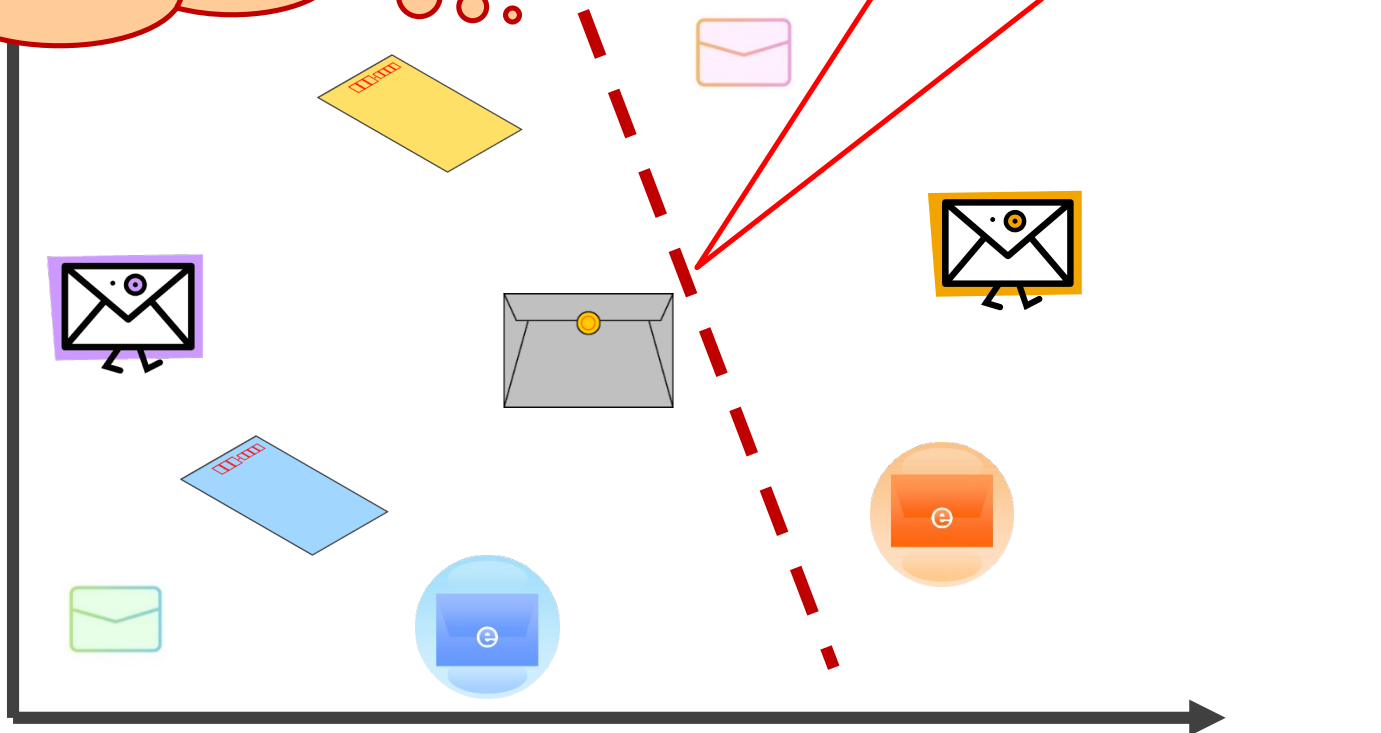
スパム!

単語Bの数

スパム分類問題

もし、こんな線を引いてしまうと...

簡単なサイドビジネスで高収入を得ませんか？...



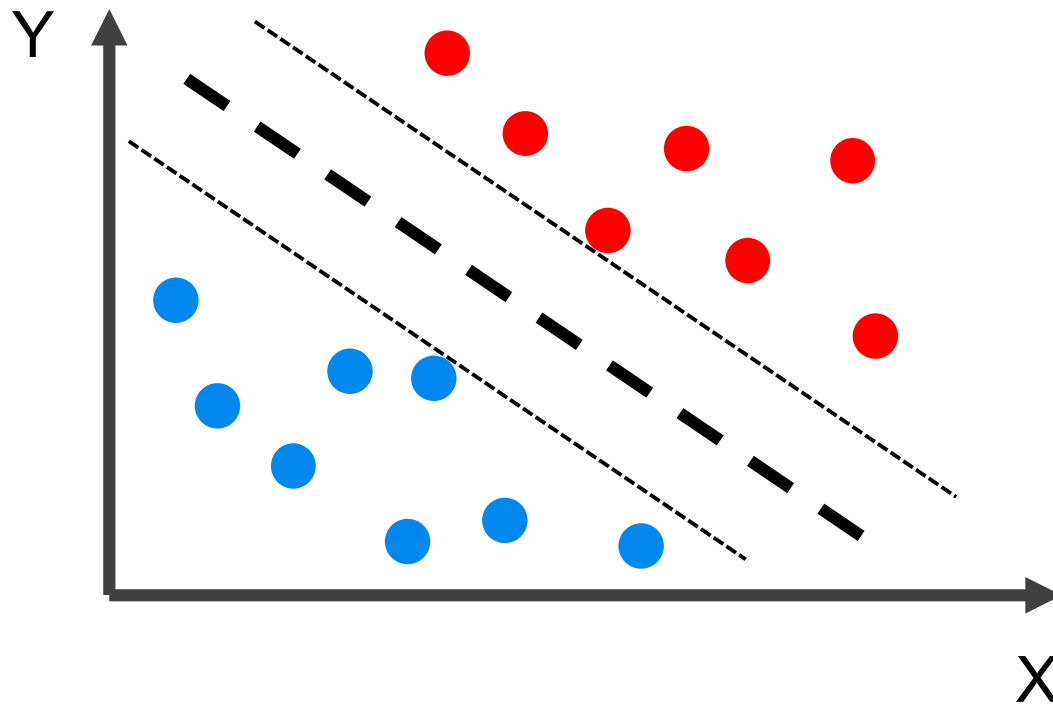
スパムじゃないですねー

← 間違った分類

単語Bの数

線の引き方 (SVM)

- 2つのグループから最も距離の離れた場所 (最大マージン) を見つけ, その真ん中に線を引く.



オンライン予測問題

- 逐次的に与えられるデータに対して、そのたびに意思決定を行う。

※ ここでいう「オンライン」とは、「逐次的」という意味で、インターネットに繋がっているという意味ではない。

オンライン勝敗予測問題

- 逐次的に与えられるデータに対して、そのたびに意思決定を行う

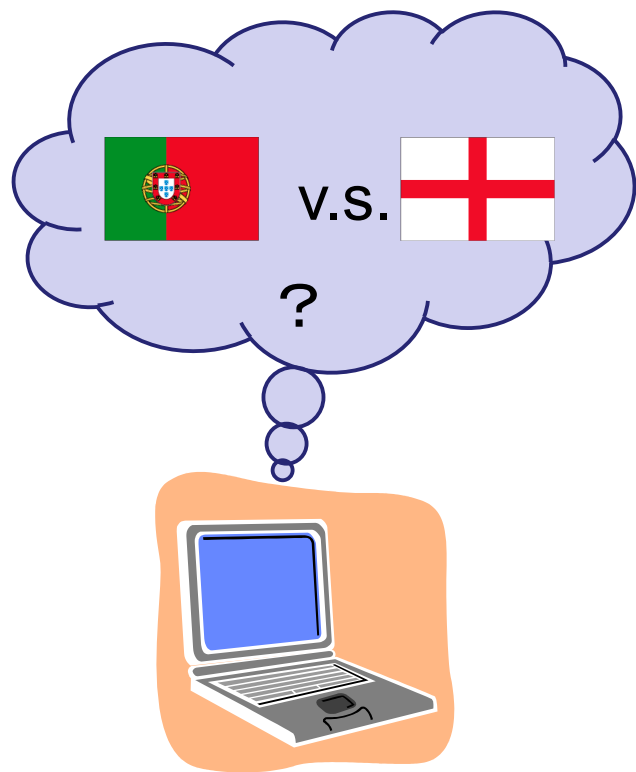
今までの対戦結果





				
		1	0	2
	3		2	1
	4	2		3
	2	3	1	

オンライン勝敗予測問題

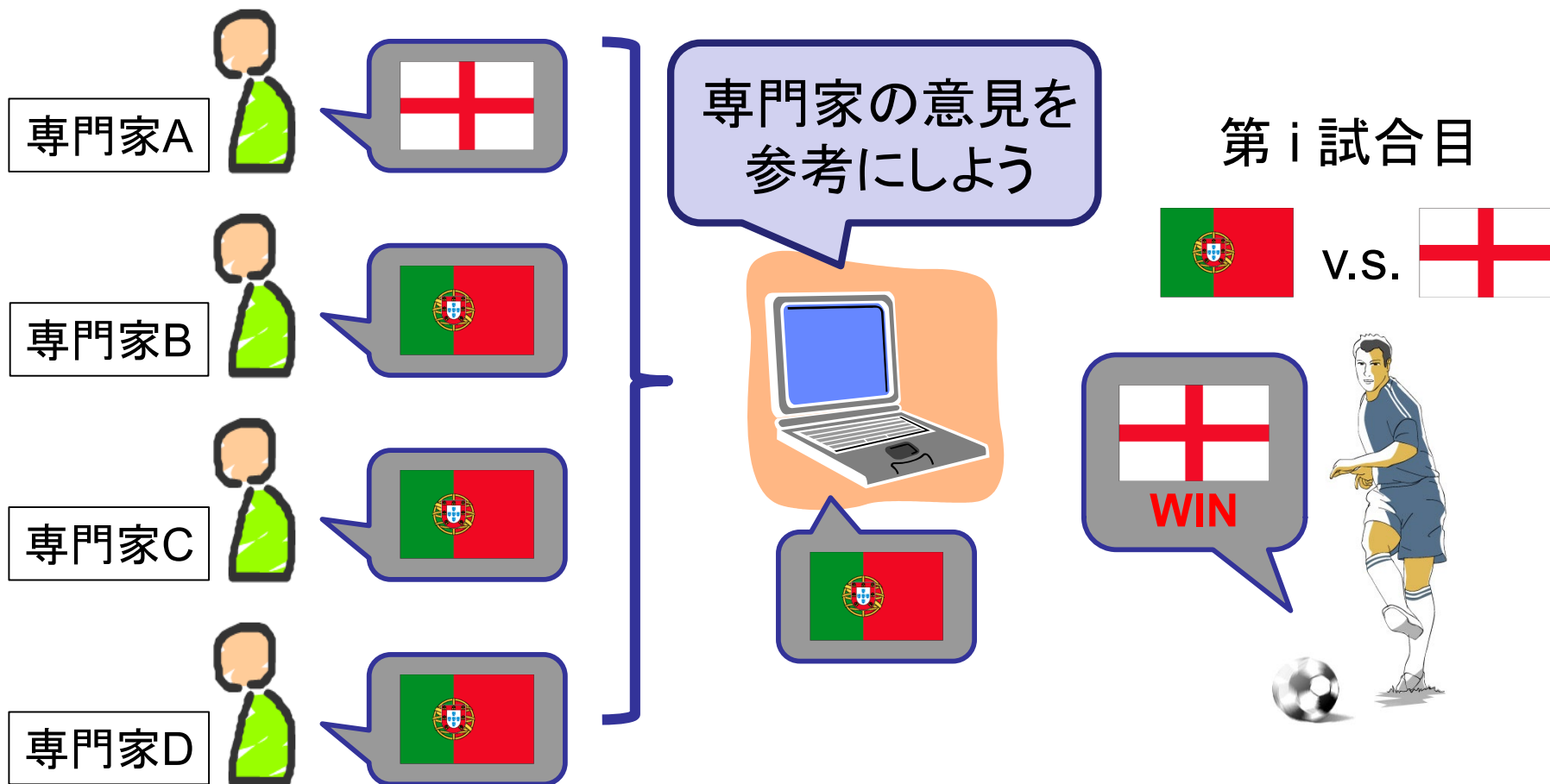
- 逐次的に与えられるデータに対して、そのたびに意思決定を行う



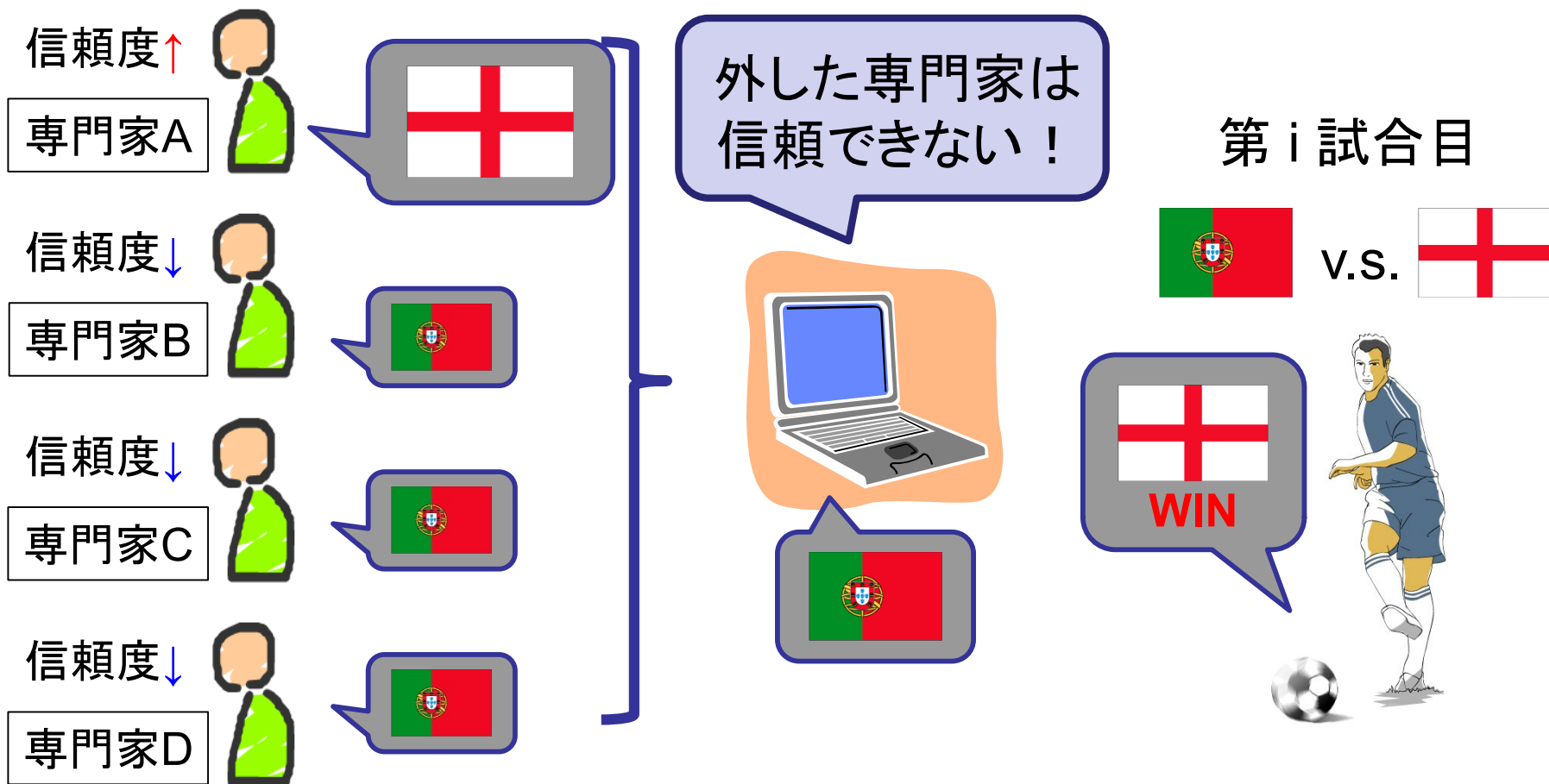
今までの対戦結果

				
		1	0	2
	3		2	1
	4	2		3
	2	3	1	

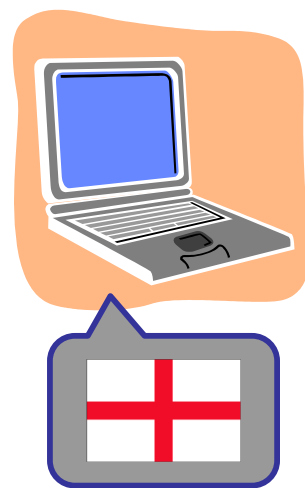
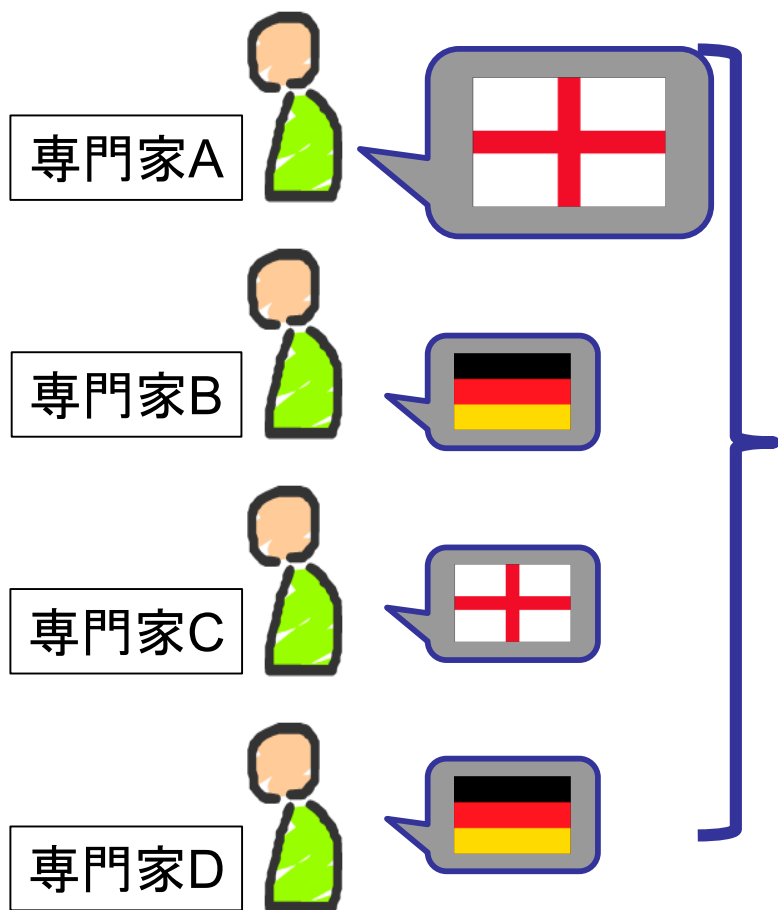
オンライン勝敗予測問題



オンライン勝敗予測問題



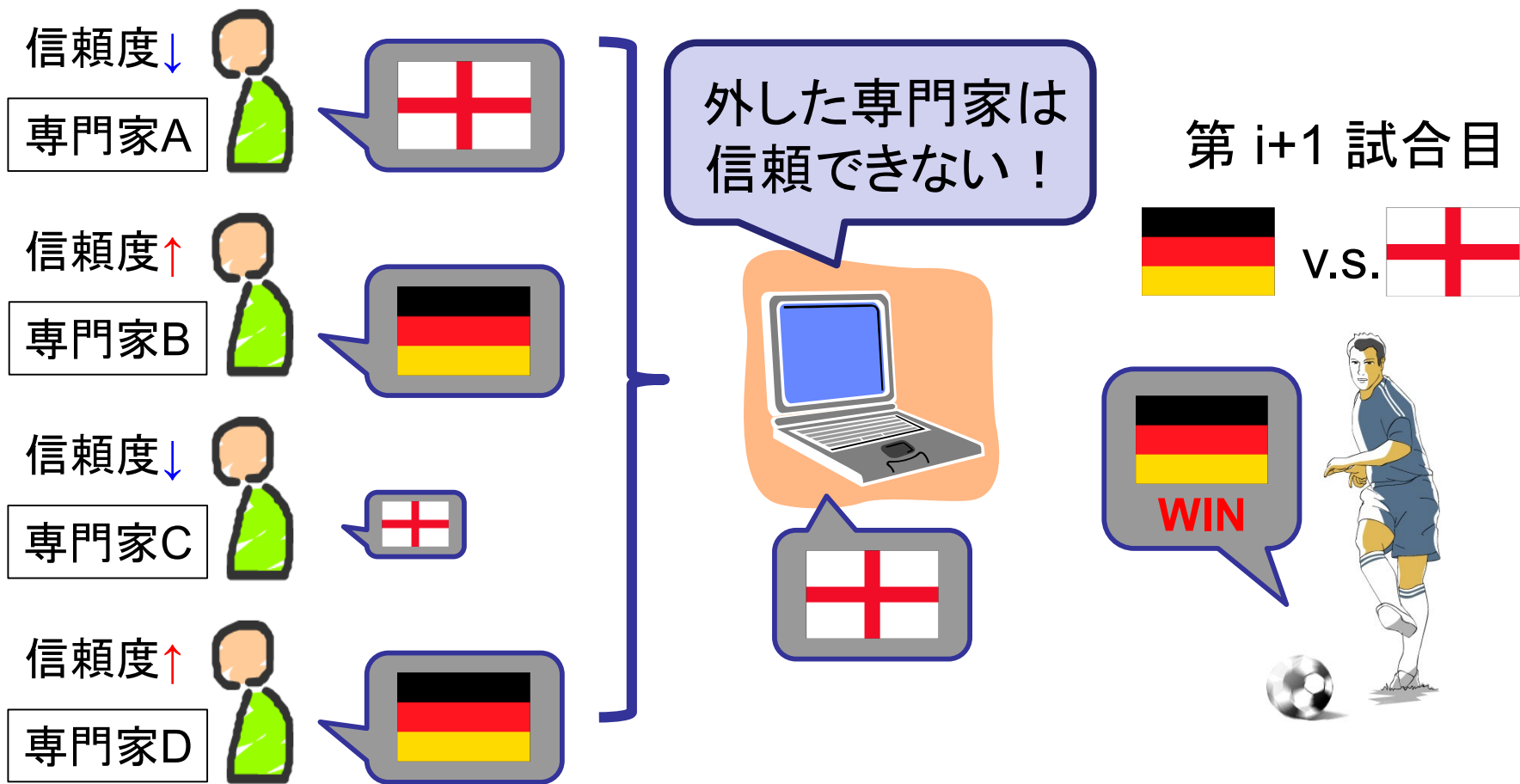
オンライン勝敗予測問題



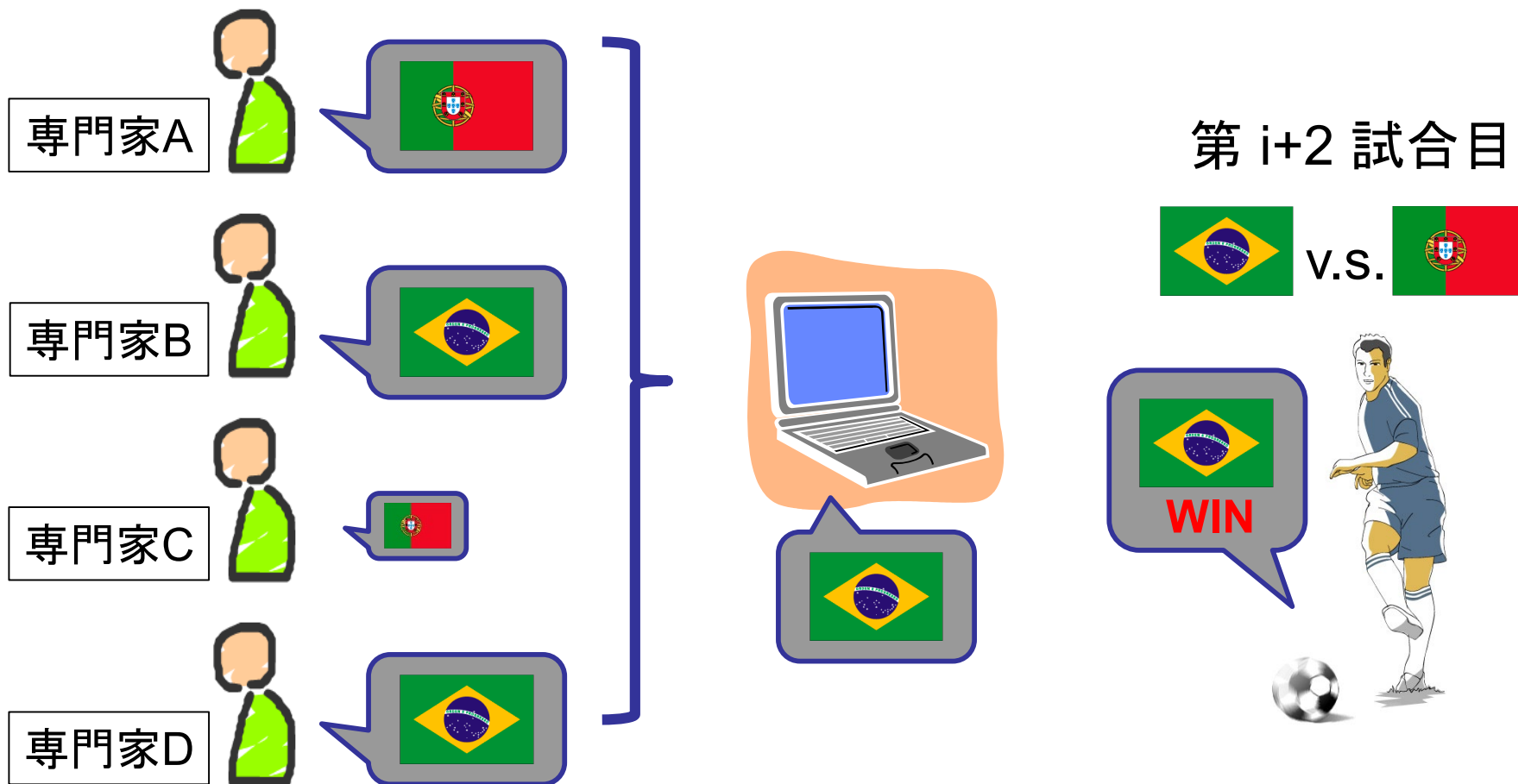
第 $i+1$ 試合目



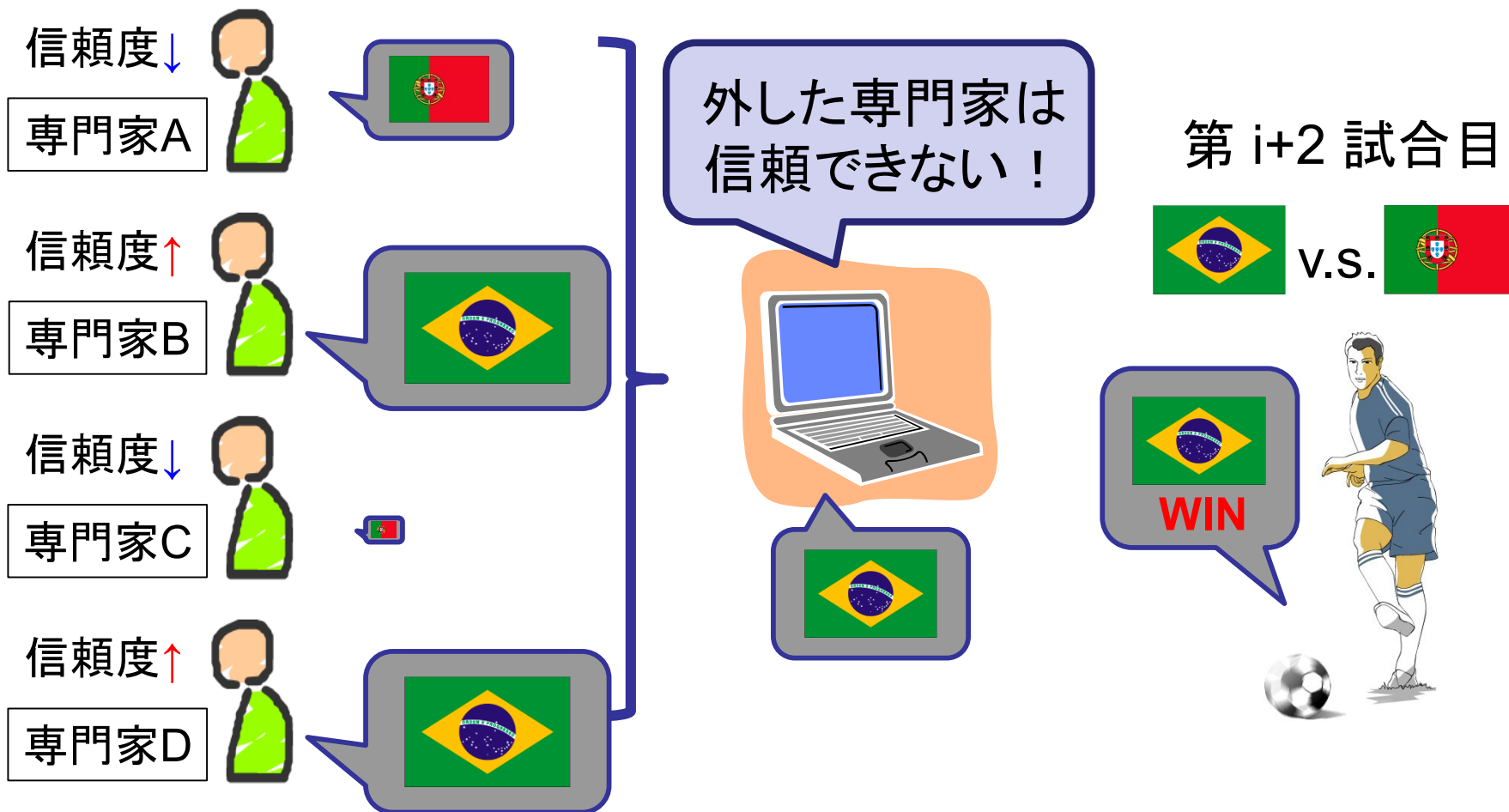
オンライン勝敗予測問題



オンライン勝敗予測問題



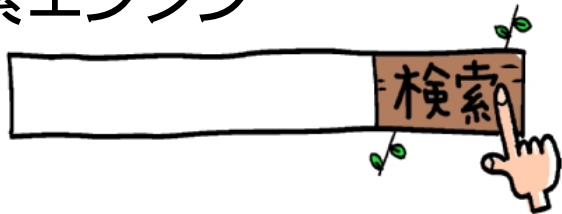
オンライン勝敗予測問題



機械学習の応用

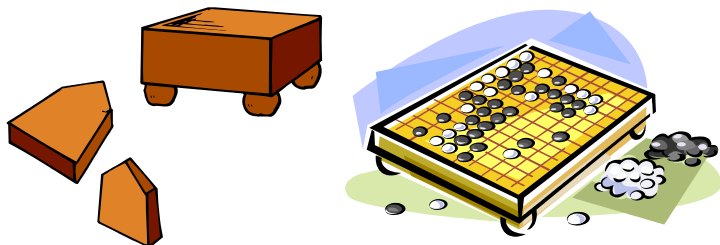
機械学習技術の応用例

・検索エンジン



・文字認識

・コンピュータ囲碁・将棋



・Web広告最適化



・株式市場の予測

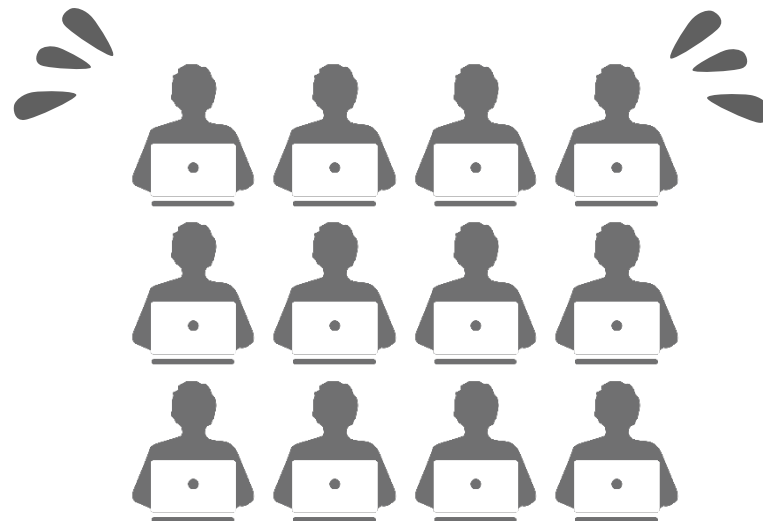
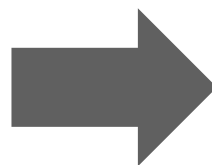


機械学習の応用 1 : 裁判の証拠レビュー支援

- 米国では, 原告・被告双方の主張立証に必要な証拠をできるだけ裁判所に提出して, 公平な裁判を行おうとするディスカバリー(証拠開示手続)という制度がある.
- 電子データに対しても, この制度が適用される.
→ 膨大なファイルから, 証拠を含むものだけを見つけだす必要がある.



機械学習の応用 1 : 裁判の証拠レビュー支援

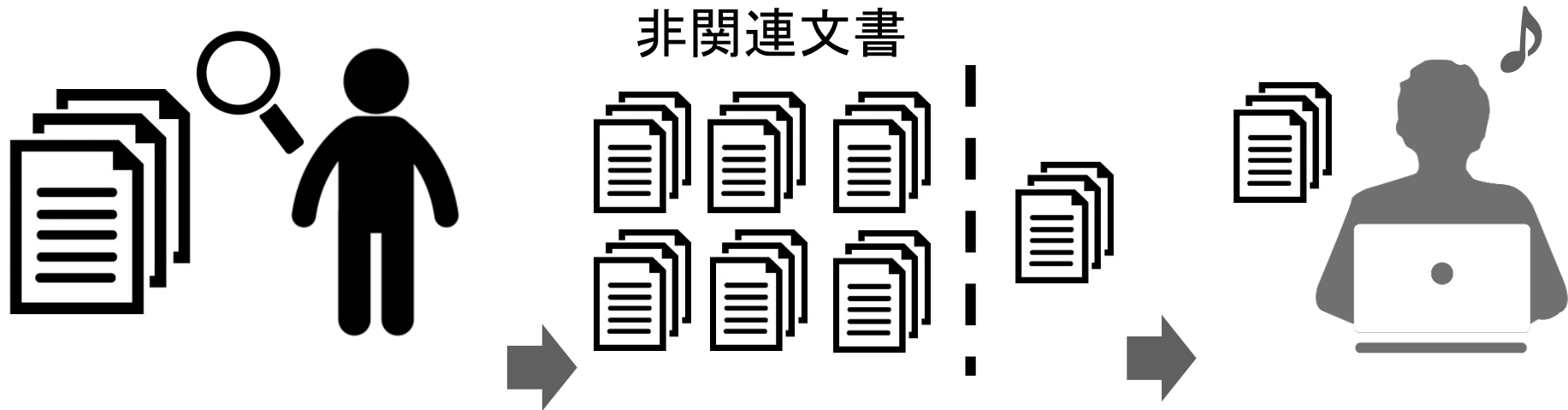


裁判の証拠を探す人は、
何が証拠になっている
かすら、わからない

証拠抽出のためには、
膨大な弁護士レビュー
が必要！

図はFRONTEO 社 HP を参考に作成

FRONTEO 社のレビュー支援システム



まず、顧客は少数の文書を閲覧し、裁判への関係性の有無を判断（訓練データの作成）

次に、コンピュータが訓練データを学習し、全レビュー対象文書を自動仕分け

最後に、関連性の高い文書だけを弁護士がレビュー

図はFRONTEO 社 HP を参考に作成

機械学習の応用2：高頻度取引

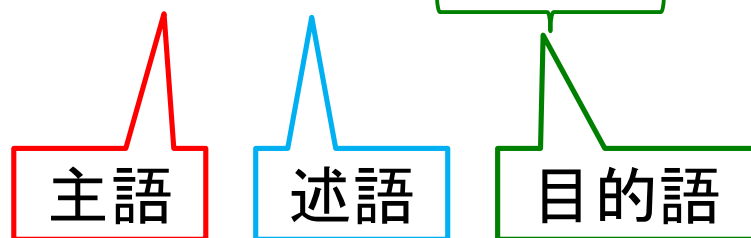
- マイクロ秒単位の極めて短い時間単位で、コンピューターによって市況を自動的に判断し、自動的に株のやり取りを実施するシステム。
- アメリカでは全取引高の50%以上が高頻度取引。
- 現在では、マイクロ秒の1000分の1のナノ秒単位の戦いになっている。
- 人間が介入する余地なし！



機械学習の応用3：自然言語処理

- 自然言語処理：人間が日常的に使っている自然言語をコンピュータに処理させる一連の技術.
- 従来の自然言語処理の手法 ⇒ 構文解析

例) I have a pen .



統計的自然言語処理

現在の主流はこれ

- 従来の方法(ルールベース自然言語処理)とはまったく異なるアプローチ.
- 対訳コーパスという, 日本語と英語が両方記載された大量の文書データから学習した情報をもとに, 英文を訳する際には

「英語のこの単語(フレーズ)は日本語のこの単語(フレーズ)に訳される確率が高い」

という感じに, 単純に当てはめていく.

機械学習の応用4：コンピュータ将棋

将棋とは

	9	8	7	6	5	4	3	2	1	
	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲

- 9×9のマスの中で、動き方8種類、40枚の駒を動かし、玉を取れば勝ち
- 相手の駒を取り、自分の持ち駒として使うことができる
- 将棋人口は約1300万人 (レジャー白書調べ)
- ニコニコ生放送カテゴリ別視聴数ランキングTOP3 (羽生善治 & ドワンゴ社長対談)

コンピュータ将棋の歴史

- 1974年 コンピュータ将棋開発開始？
(探索と評価関数の試行錯誤)
- 2005年 「激指」がアマ竜王戦ベスト16
- 2006年 「Bonanza」登場
(評価関数に機械学習を導入)
- 2010年 「あから2010」が清水女流王位(当時)に勝利
- 2012年 「ボンクラーズ」が米長永世名人(引退棋士)に勝利
- 2012年 電王戦にてコンピュータが現役プロ棋士チームに
勝ち越し

以後コンピュータ将棋の優勢が続く(トップ棋士も敗れる)

強いコンピュータ将棋ソフトを作るには

なるべく先まで読み, 正確に評価する

先読み(探索)

- コンピュータの計算力向上で近年大幅にパワーアップ
 - ・ トップソフトは数百万～数億手/秒 読む
- 従来の候補手を絞った探索から,
全幅探索(網羅的に手を読む)が可能に

局面評価(評価関数)

- 機械学習技術によるパワーアップ
- 従来の手作業によるチューニングから,
数億もの評価項目の自動調整が可能に

先読み



角得&馬ができた！？

先読み



角得&馬ができた！？



次の相手の手を
考えれば, 実は互角

ミニマックス法

- 対戦相手がいるゲームでは、相手が次に打ってくる手を考慮して、自分の手を決める必要がある。
- 各盤面に点数をつける（点数の付け方は後述）。
- 自分は、自分の点数を最大（Max）にする手を打つ。
- 相手は、自分の点数を最小（Min）にする手を打つ。
- この考え方に基づいて、次に自分の打つ手を決める方法を **ミニマックス法** という。

評価関数（評価基準）

- 評価項目：どんな特徴を盤面の点数に反映させるべきか？
- 値の調整：各特徴に、どのくらいの点数を設定すればよいか？

例：駒に点数をつけて評価

飛	角	金	銀	桂	香	歩	龍	馬	成銀	成桂	成香	と
170	140	100	80	60	50	20	185	160	91	82	78	65

代表的な評価項目

- 駒の価値
- 玉に対する金の位置
- 玉の危険度(玉周辺の敵駒の利き)
- 各駒の働き(利き, 敵, 味方の駒との位置関係)
 - 桂馬が跳べるか
 - 歩が進めるか
 - ひもがついているか

などなど専門的で複雑なものがたくさん

例：どっちの金が価値が高い？

同じ「金」でも「玉」に近いほど価値が高い。

→ このような評価項目が膨大にある。

9	8	7	6	5	4	3	2	1	
									一
									二
									三
									四
									五
									六
									七
	金								八
玉									九

9	8	7	6	5	4	3	2	1	
									一
									二
									三
									四
									五
									六
									七
	金								八
								玉	九

代表的な評価項目

- 駒の価値
- 玉に対する金の位置
- 玉の危険度(玉周辺の敵駒の利き)
- 各駒の働き(利き, 敵, 味方の駒との位置関係)
 - 桂馬が跳べるか
 - 歩が進めるか
 - ひもがついているか

などなど専門的で複雑なものがたくさん

組み合わせを考慮すると**数億**に及ぶ

機械学習の応用5：クイズ

- 2011年, IBM が開発した質問応答システム「ワトソン」がアメリカのクイズ番組 Jeopardy! で人間のクイズチャンピオンと対戦して勝利.
- オンライン百科事典「ウィキペディア」から学習した知識を利用して, クイズに回答する.



WIKIPEDIA
The Free Encyclopedia

- 一方で, ワトソンが回答を出すプロセスは, 人間の思考方法とはかなり異なる.

ワトソンの応用

- 人材マッチング（フォーラムエンジニアリング）
 - ワトソンが個人の興味, 関心, 性格などを数値化して把握し, マッチングスコアを算出する.
2016年4月から実際に運用されている.
- 医療アドバイス（東大医科研）
 - がんに関連する約2千万件の論文をワトソンが学習.
 - 60代の女性患者の白血病が治療が難しい特殊なタイプであるとの分析を10分で行い, 治療法の変更を提案.
- 他にもいろいろ

深層学習

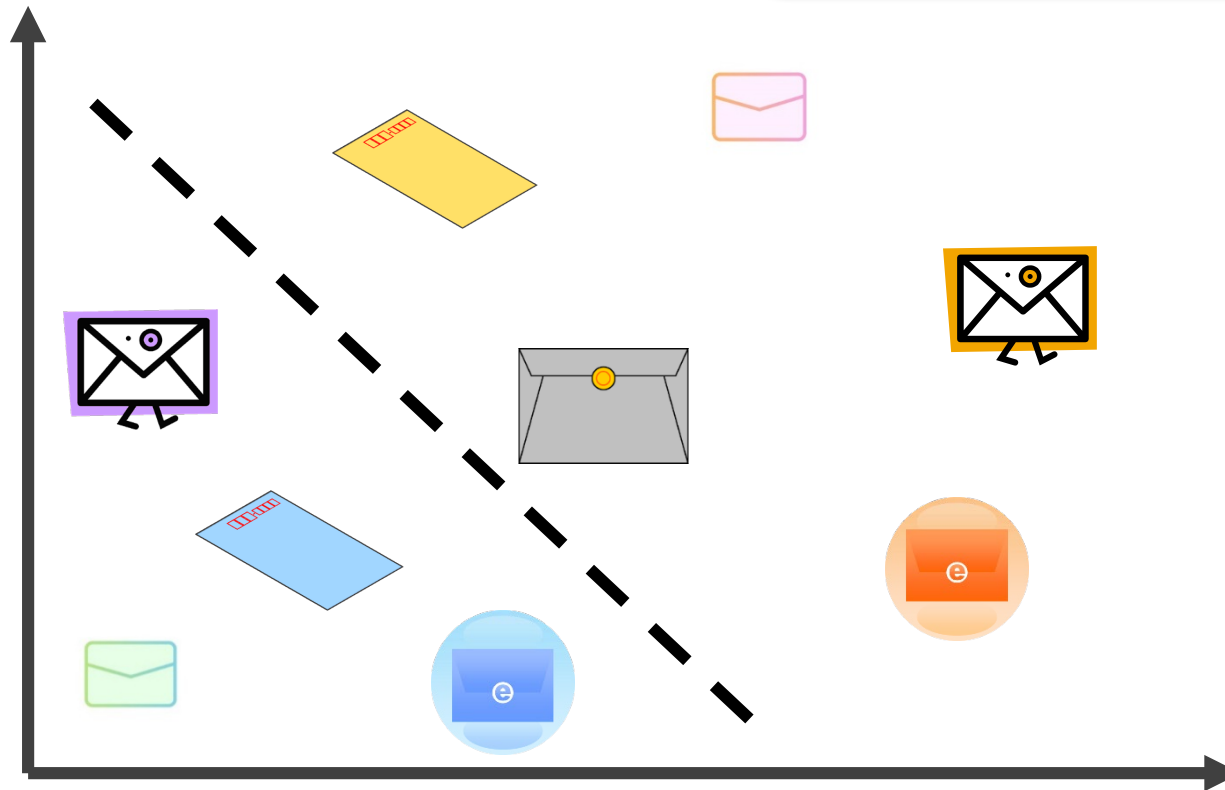
評価項目の選定の難しさ

- 結局、機械学習技術の性能、ひいては人工知能の性能は、対象をどんな項目で評価するかに大きく依存する。

スパム分類問題

- どんな単語を選ぶべきか？
- 何種類の単語を使うべきか？
- 単語の相関関係は？

単語Aの数



単語Bの数

コンピュータ将棋

コンピュータ将棋の代表的な評価項目

- 駒の価値
- 玉に対する金の位置
- 玉の危険度(玉周辺の敵駒の利き)
- 各駒の働き(利き, 敵, 味方の駒との位置関係)
 - 桂馬が跳べるか
 - 歩が進めるか
 - ひもがついているか

全部使うわけにはいかない
➤ 計算時間がかかりすぎる
➤ 不要な項目もたくさんある

組み合わせを考慮すると**数億**に及ぶ

職人技

- 既存の機械学習の手法では, 評価項目の選定と値の設定は人間が手作業で行っていた.
- 職人技によるチューニングの世界.



飛	角	金	銀	桂	香	歩	龍	馬	成銀	成桂	成香	と
170	140	100	80	60	50	20	185	160	91	82	78	65

職人技

- 既存の機械学習の手法では、評価項目の選定と値の設定は人間が手作業で行っていた。
- 職人技によるチューニングの世界。



この辛い作業を自動化したい...

飛	角	金	銀	桂	香	歩	龍	馬	成銀	成桂	成香	と
170	140	90	80	60	50	25	185	160	91	82	78	65

深層学習 (Deep Learning)

- データから, コンピュータが自動的に評価項目を作り出す手法.
 - 人間による評価項目の選定, 数値の設定(職人技)を必要としない.
- 現在の人工知能ブームは, 深層学習の発達・普及によるところが大変大きい.

人工知能の現状

- 現在の人工知能技術は、以下の2つの要素技術を組み合わせて作られている。

➤ 機械学習



画像 http://www.cs.toronto.edu/~urtasun/courses/CSC411_Fall16/CSC411_Fall16.html

➤ ビッグデータ解析

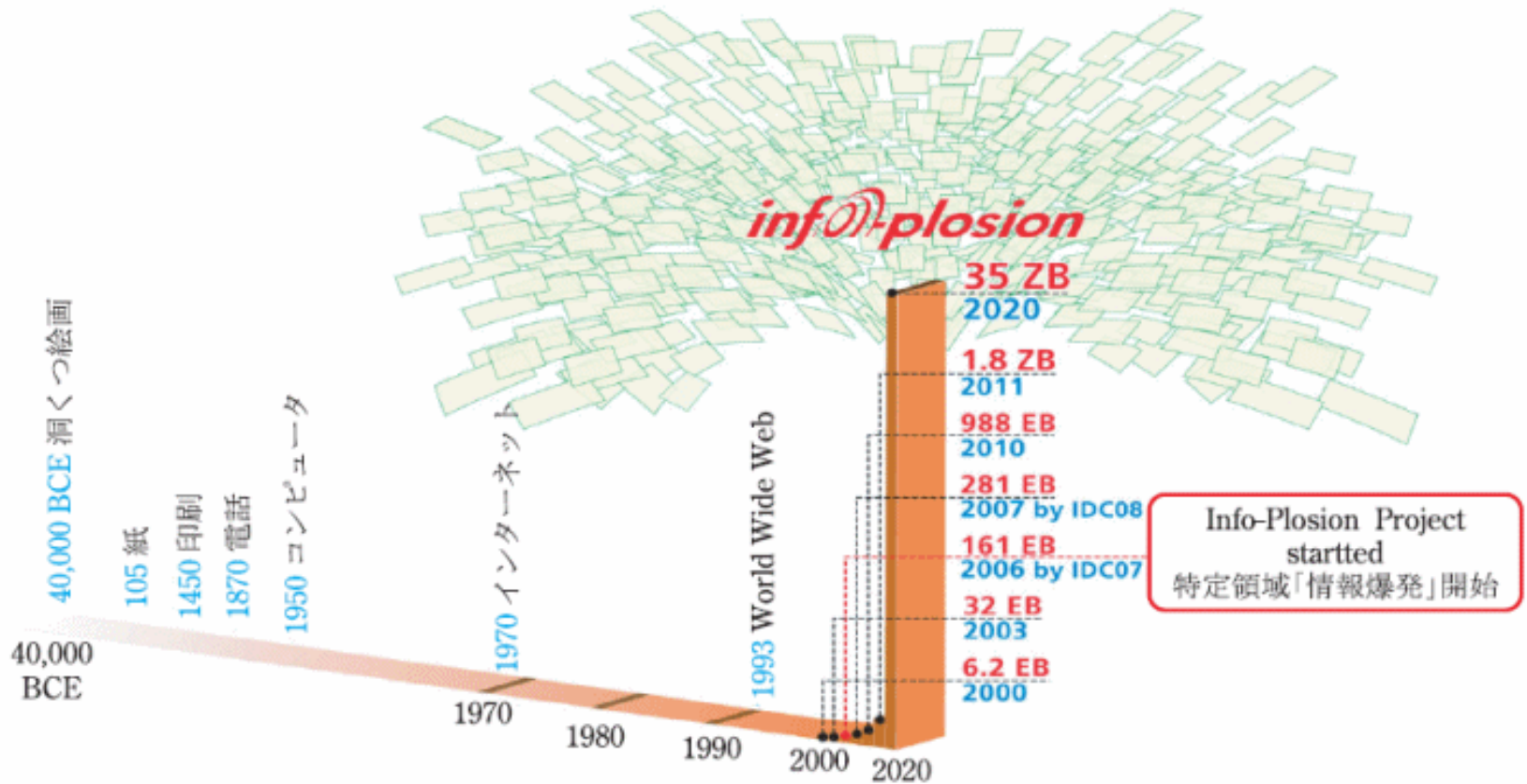


画像 <http://www.trickymindsolution.com/big-data-analysis-services.html>

ビッグデータ解析

情報爆発時代

- 電子データの量が「爆発的に」増加している！



喜連川優「情報爆発のこれまでとこれから」,電子情報通信学会誌, Vol.94, No8, 2011

ビッグデータ

毎日、大量のデータがインターネット上に生成され、記録されている。

- Google: 60 億件の検索
- Facebook: 43 億メッセージ
- Twitter: 5 億ツイート
- YouTube: 400 万時間分の動画

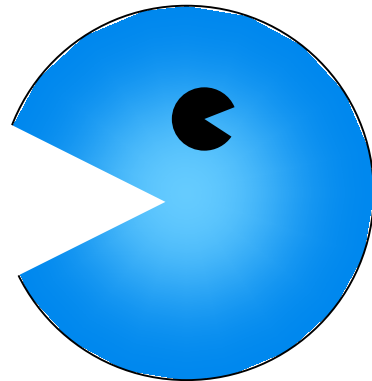
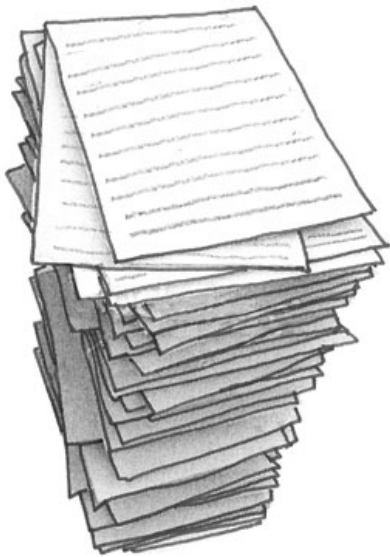


データは <https://www.gwava.com/blog/internet-data-created-daily> より

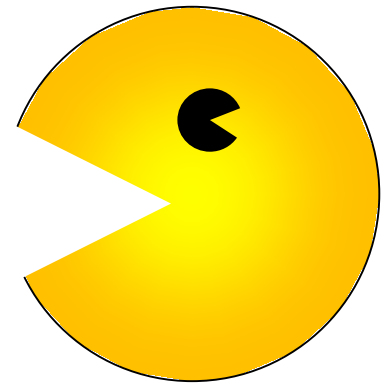
私の研究

【圧縮文字列情報処理】

- デジタルデータを**文字列**（記号の連鎖）として捉え、
- データを**圧縮**したまま処理する技術。



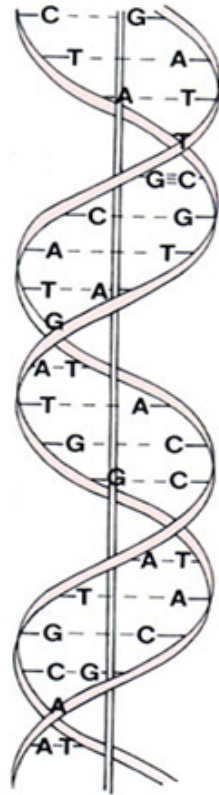
既存技術



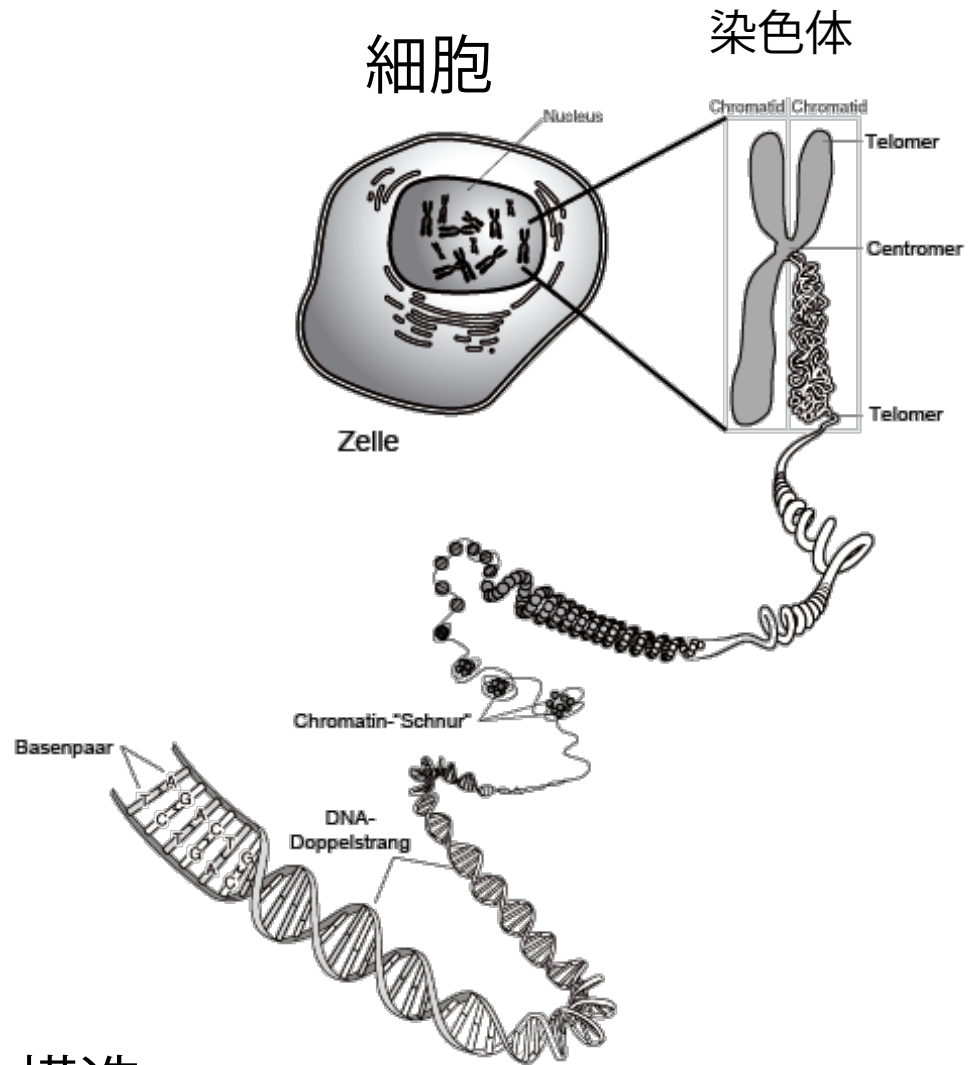
開発した技術

万物は文字列である

例1) DNA配列



A, C, G, T からなる
2つの文字列のらせん構造



万物は文字列である

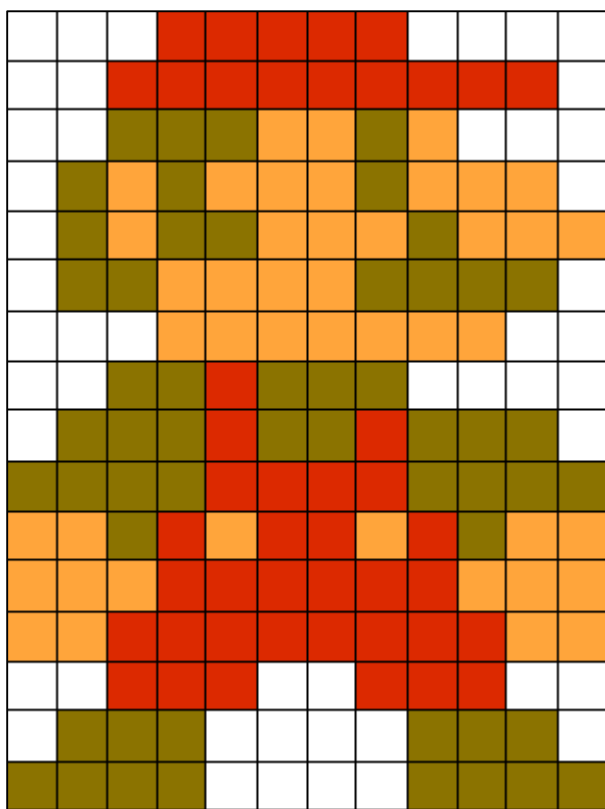
例2) 音楽データ



音符と休符の列

万物は文字列である

例3) 画像データ

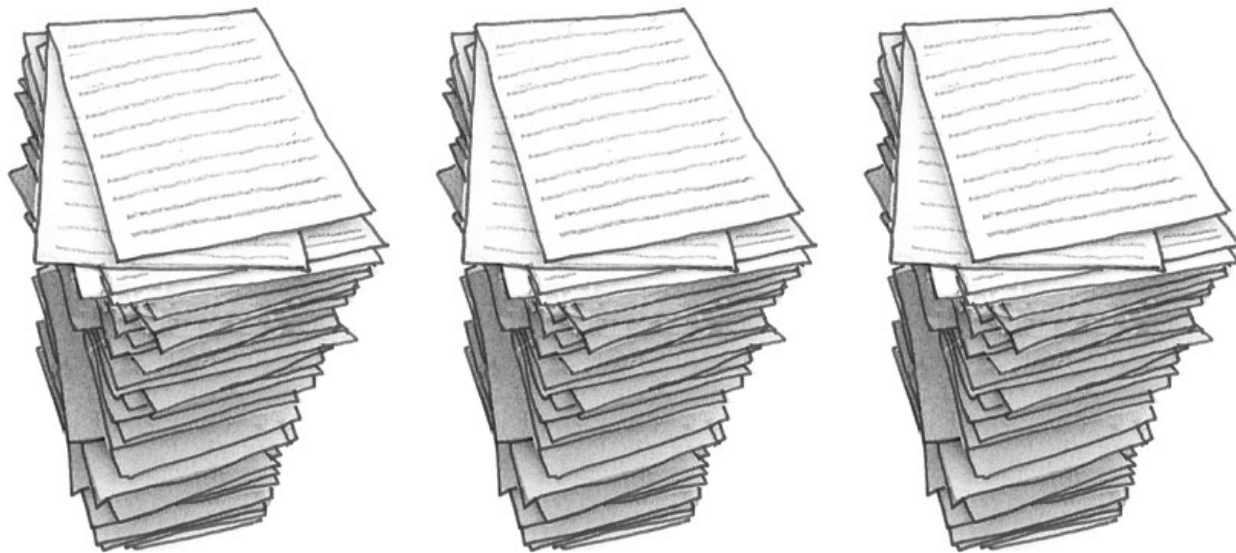


■ □ ■ ■ の二次元列

色を記号だとみなせば
画像も文字列である

コンピュータ上のデータ = 文字列

- ◆ コンピュータ上のデータの多くは文字列とみなせる.
- ◆ 大量に溢れる文字列データをいかに効率よく処理するかがカギ.



データ圧縮

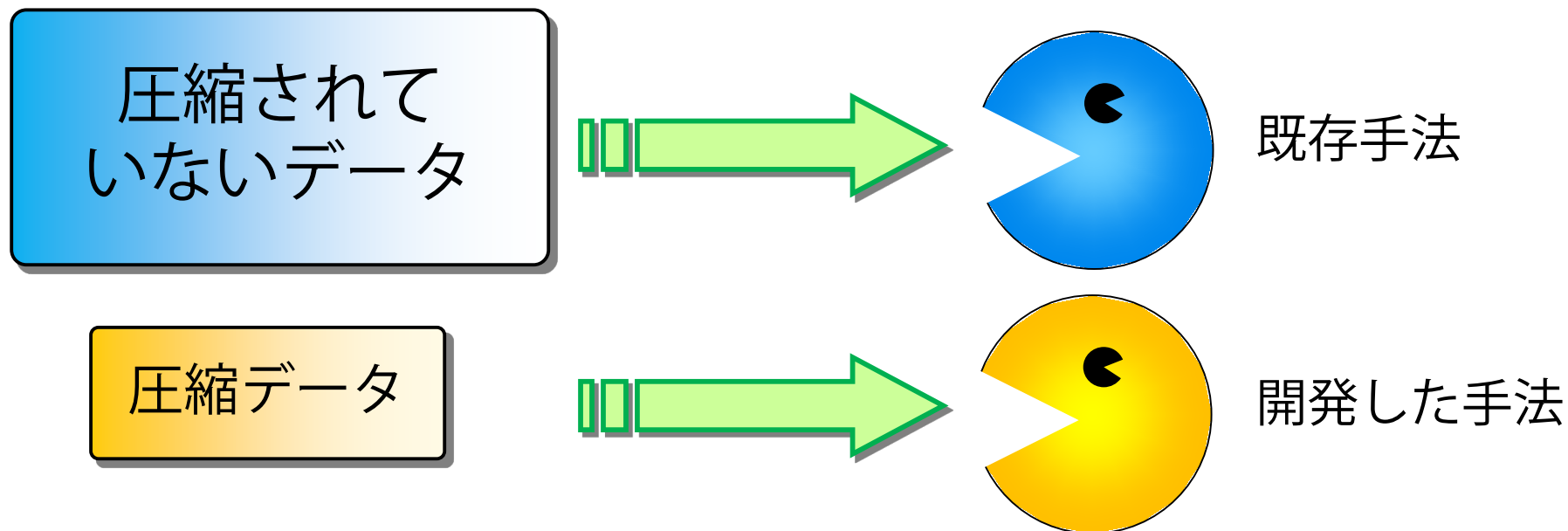
文字列データを圧縮して表現すれば、データの「無駄」を削減できる。

例1) すももももももものうち
= すも⁸のうち

例2) にわにわにわにわとりがいる
= (にわ)⁴とりがいる

研究成果

- 圧縮された文字列データ上で情報検索を高速に行う手法を開発した。
 - 大量のデータも圧縮すれば小さくなる
 - 圧縮することで、情報検索の速度も向上！



人間と人工知能

人工知能が人間の職を代替する？

- すでに20世紀の時点で、工場等における単純作業や力仕事の自動化（機械化）はかなり進んでいた。
- 人工知能の発達によって、「より知的」と考えられていた職業も、人工知能が代替する可能性がある。

人工知能が人間の職を代替する？

- 人工知能によって技術的に代替できる可能性が高い職業リスト(一部)

- スーパー店員
- 測量士
- 宝くじ販売人
- 電車運転士
- 事務員
- ビル清掃員
- 銀行窓口係
- 出荷・発送係員
- ...

- 型にはまった仕事が多い
- イレギュラーな要素が少ない

※ リストは野村総研の研究資料より抜粋

人工知能が人間の職を代替する？

- 人工知能によって技術的に代替できる可能性が低い職業リスト(一部)

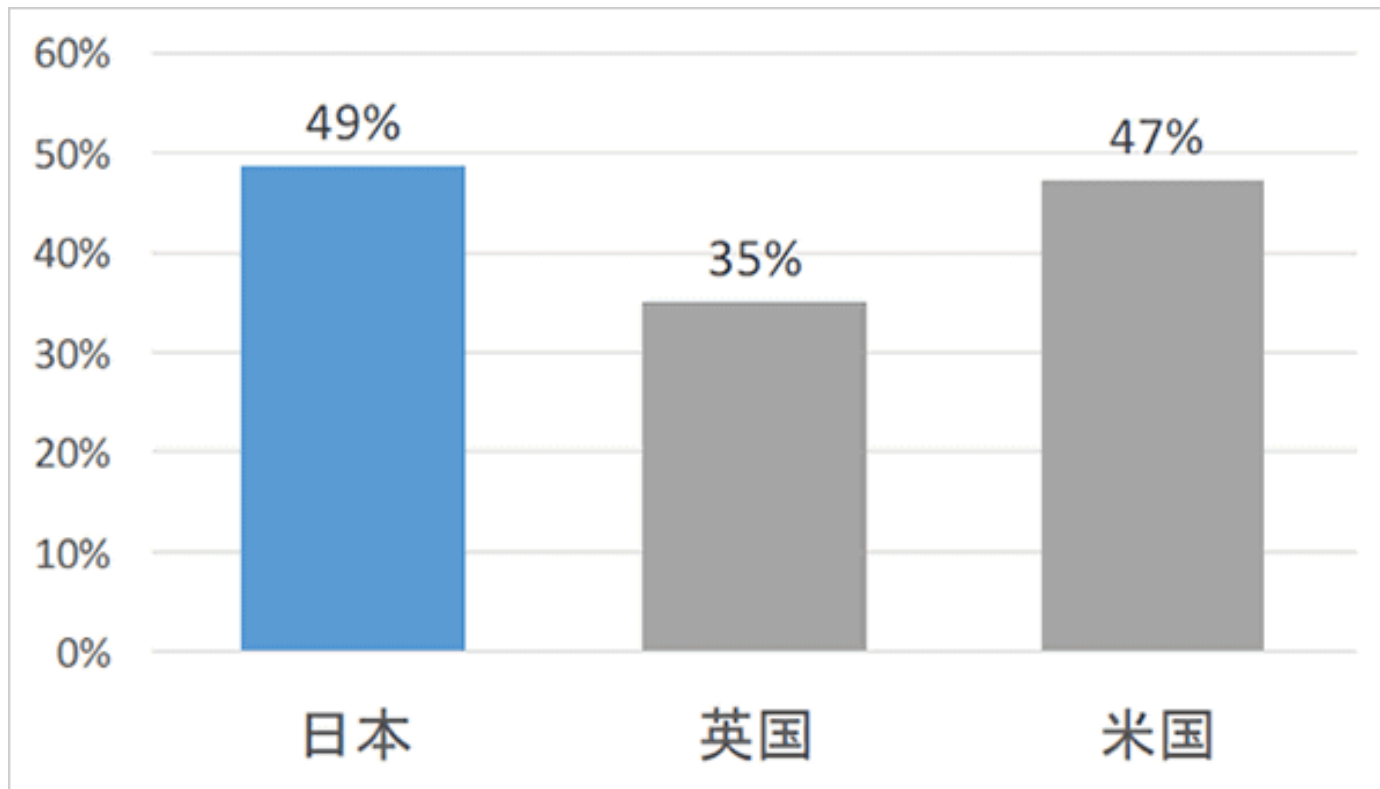
- 映画監督
- 漫画家
- 教員
- 保育士
- ミュージシャン
- 医師
- 美容師
- 評論家
- ...

- 創造性が求められる仕事
- 社会的知性(単純ではない対人業務)が求められる仕事

※ リストは野村総研の研究資料より抜粋

人工知能が人間の職を代替する？

- 人工知能やロボットによる代替可能性が高い職業に就く人口の割合



※ 図は野村総研の研究資料より

人工知能は万能ではない！

- そもそも、現在主流になっている人工知能技術は「なんとなく上手くいっている」だけで、実は、なぜ上手くいっているかわかっていない。
 - 結果の精度・質に対する保証がない。
- すべてを鵜呑みにすると、大きな落とし穴があるかもしれない？

まとめ

- 現在の人工知能技術は
 - 機械学習
 - ビッグデータ処理という2つの要素技術に支えられている.
- 深層学習（Deep Learning）の登場.
- 人工知能は万能ではない.
- 人工知能を上手く利活用し、共存することが重要！



参考図書

- 「人工知能は人間を超えるか
ー ディープラーニングの先にあるもの ー」
松尾 豊 (著)

