

人工知能に関する技術動向と 産業分野への利用可能性

東京大学 松尾 豊

もくじ

- 人工知能、ディープラーニングとは何か
- 今後の技術の進展
- それにともなう産業の可能性、やるべきこと

人工知能とは？

- 人のようにしゃべる
- 人のように考える
- 感情をもつ、意思を持つ
- 意識をもつ
- たくさんデータを使う
- どんどん覚えていく
- 決して間違えない
- ...



普通のひとが考える

AI

そんなに簡単ではない。
過去に、さまざまな研究や議論が行われてきた。

人工知能をめぐる動向

- 第1次AIブーム(1956～1960年代):探索・推論の時代

- ダートマスワークショップ(1956)
 - 人工知能(Artificial Intelligence)という言葉が決まる
 - 世界最初のコンピュータENIAC(1946)のわずか10年後
- 数学の定理証明、チェスを指す人工知能等

考えるのが早い人工知能

- ...冬の時代

- 第2次AIブーム(1980年代):知識の時代

- エキスパートシステム
- 医療診断、有機化合物の特定、...
- 第5世代コンピュータプロジェクト:通産省が570億円

ものしりな人工知能

- ...冬の時代

- 第3次AIブーム(2013年～):機械学習・ディープラーニングの時代

- ウェブとビッグデータの発展
- 計算機能力の向上

データから学習する人工知能

- 多くのイノベーションは、初期の10年～20年に生まれている。
- 過去の技術の蓄積があり、ビッグデータに対して活用できる。
例えば、推論、制約充足、対話システム、検索・情報推薦、画像・音声認識、ユーザモデリングなど

将棋電王戦



IBM ワトソン



ディープラーニング(2007-)

ILSVRCでの圧勝(2012)
Googleの猫認識(2012)
ディープマインドの買収(2013)
FB/Baiduの研究所(2013)

機械学習

ウェブ・ビッグデータ

車・ロボット
への活用

自動運転
Pepper

統計的自然言語処理
(機械翻訳など)

検索エンジンへの活用

MYCIN(医療診断)
DENDRAL

エキスパート
システム

オントロジー

タスクオントロジー

ワトソン(2011)

LOD (Linked Open Data)

Eliza

対話システムの研究

Caloプロジェクト

Siri(2012)

bot

探索
迷路・パズル

プランニング
STRIPS

チェス(1997)
Deep Blue

将棋(2012-)
電王戦

囲碁

1956

1970

1980

1995

2010

2015

第一次AIブーム
(推論・探索)

第二次AIブーム
(知識表現)

第三次AIブーム
(機械学習・ディープラーニング)

これまでの人工知能の壁＝特徴抽出の壁

- 難しい問題1：機械学習における素性設計 (Feature engineering)
 - － 機械学習において、素性(特徴量)の設計が難しかった。
 - － 人間が対象をよく観察して設計するしかなかった。
- 難しい問題2：フレーム問題
 - － 人間が知識を記述することで、人工知能を動作させる。
 - － そのときに、いくら知識を書いても、うまく例外に対応できない。
- 難しい問題3：シンボルグラウンディング問題
 - － シマウマがシマのある馬だと、計算機が理解することができない。
 - － シンボル(記号)がそれが指すものと接続(グラウンド)しておらず、シンボルの操作ができない。



結局のところ、いままでの人工知能は、

人間が現実世界の対象物を観察し、「どこに注目」するか(特徴量の抽出)を見ぬいて、モデルの構築を行っていた。

その後の処理は自動で行うことができたが、モデル化の最初の部分に人間が大きく介在していた。それが、唯一にして最大の問題であった。

ディープラーニングとは

- 人工知能における50年来のブレークスルー
 - 特徴量の抽出ができるようになり始めている
- 多層の(深い)ニューラルネットワーク
 - 2006年ごろのHinton(トロント大)らの研究が契機
 - 2012年の画像認識のコンペティションで躍進
- 「特徴量の抽出」自体を計算機が行う技術。
 - オートエンコーダと呼ばれる仕組みを用いる。主成分分析を、非線形・多層にしたもの。
 - 画像認識の場合はConvolutional Neural Network(畳み込みニューラルネットワーク)という仕組みが一般的。
 - 実際に精度があがったのは、ReLUという関数の利用、重みの共有、ドロップアウトなどさまざまな要因が。
 - 計算量が非常に大きく、GPUを数個から数十個用いて大規模な計算を行う。
- 2013年から各国で投資合戦が激化
 - 米国Google, Facebook、中国Baiduなどが、それぞれ数百億～1000億円の投資
 - 米国や欧州の政府投資も年間300-400億円規模

人工の神経回路、威力増す

「ディープラーニング」と呼ぶ人工知能技術が高い関心を集めている。画像や音声の認識精度が大幅に高まるため、米グーグルなどが研究に参入。経済動向の予測や新薬開発などにも威力を発揮する可能性がある。

ここ1～2年、世界中の人工知能の研究者から大きな注目を浴びている技術がある。コンピューターに人間と同じように経験に基づいた行動をさせる機械学習の一種で、「ディープラーニング」と呼ばれる新手法だ。

インターネット社会を支える画像認

識や音声認識、新薬開発に役立つ化合物の活性予測——。こうした技術の精度を競うコンテストで、ディープラーニングが過去の記録を大幅に塗り替え、次々と優勝を果たしている。「これほど飛躍的に精度が向上するとは信じられない」「まさに衝撃的な

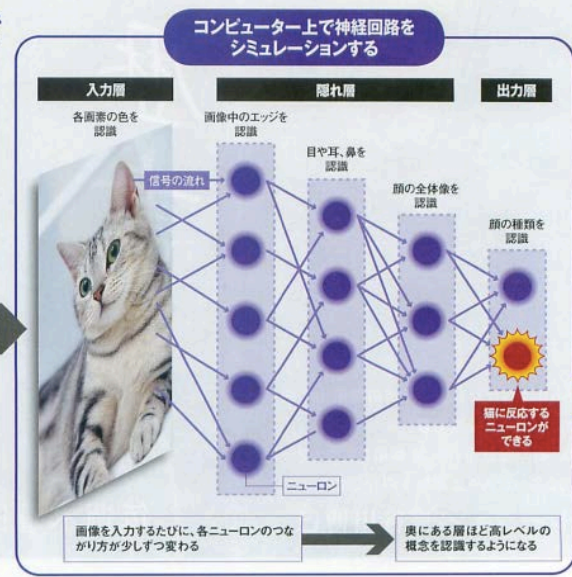
結果だ」。専門家からは、口々に驚きの声が上がっている。

ディープラーニングは、人の神経回路をコンピューター上で模倣する「ニューラルネットワーク」という技術を発展させたものだ。

人の脳は、画像からそこに映るモノ

ディープラーニングで画像を認識する流れ

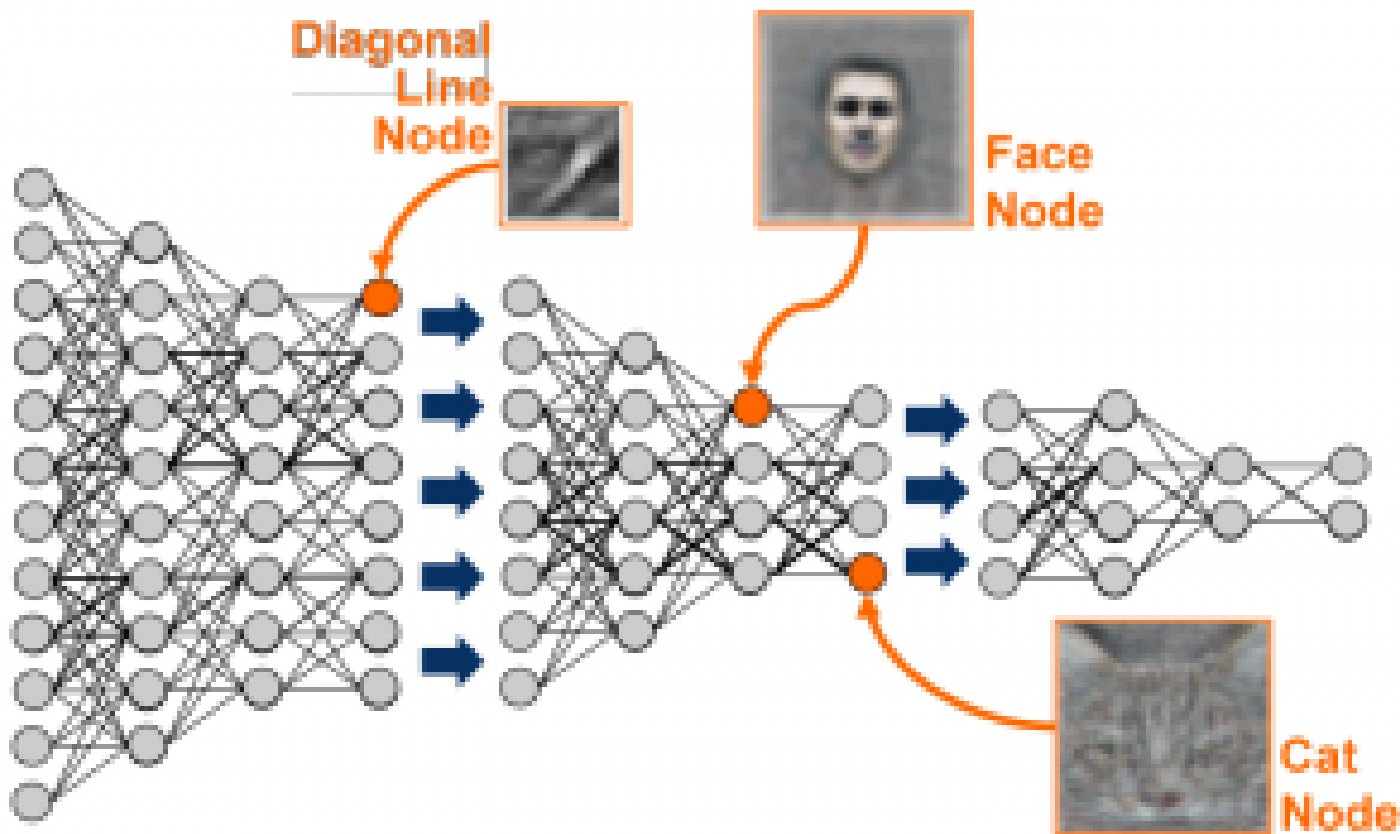
ディープラーニングでは、コンピューター上に人間の脳と同じような多層の神経回路を作製。大量の画像や文字情報を入力してトレーニング(訓練)すると、そこに含まれる高度な概念が自然に引き出される。米グーグルの研究では、出力層のニューロンが、「猫」を認識して強く反応するようになった(右は簡略化した仕組みの図)



知能の仕組みを根本から解き明かす可能性がある。

日経ビジネス2013年4月15日号

Googleの猫(2012)



- ・YouTubeから取ってきた大量の画像をニューラルネットワークに学習させることで、下位の層のニューロンには線や点といった単純な模様(特徴量)が、上位の層には、人の顔や猫といったより複雑な模様(高次の特徴量)が学習される。
- ・人間の視神経のモデルとして知られているものと極めて近い。

ディープラーニングの実績(2012)

- 画像認識の精度が急激に向上し、関係者に衝撃が走った
- ILSVRC2012 : Large Scale Visual Recognition Challenge 2012

ディープ
ラーニング

長年の
特徴量設計
の工夫

Team name	Error	Description
SuperVision	15.315%	Using extra training data from ImageNet Fall 2011 release
SuperVision	16.422%	Using only supplied training data
ISI	26.602%	Weighted sum of scores from classifiers using each FC
ISI	26.646%	Naïve sum of scores from classifiers using each FV
ISI	26.952%	Naïve sum of scores from each classifier with SIFT+FV, LBP+FV, GIST+FV and CSIFT+FV, respectively
OXFORD_VGG	26.979%	Mixed selection from High-Level SVM scores and Baseline Scores, decision is performed by looking at the validation performance.
...	...	...

「ケタ」が違う

エラー率の変化: 2012年以降

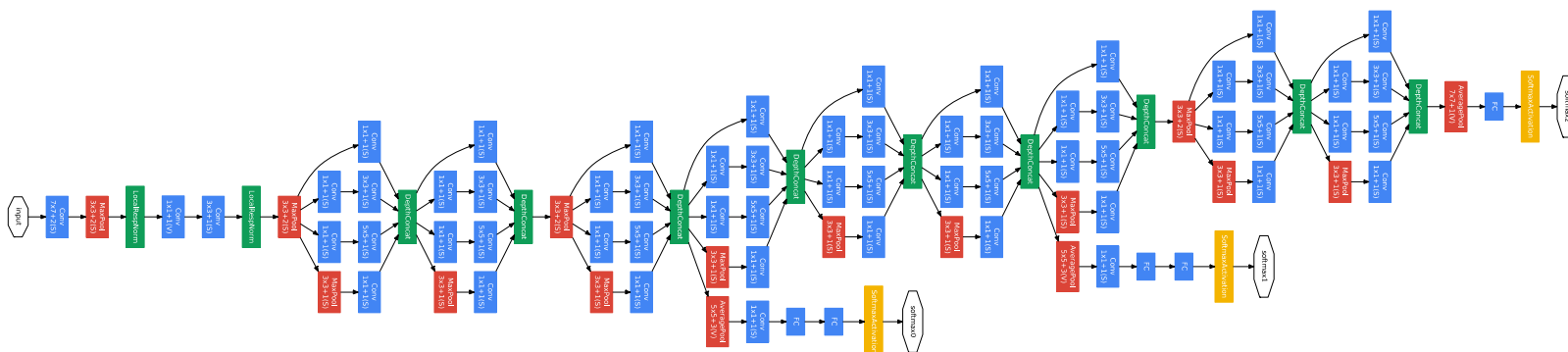
		Error
Before ディープ ラーニング	Imagenet 2011 winner (not CNN)	25.7%
	Imagenet 2012 winner	16.4% (Krizhevsky et al.)
	Imagenet 2013 winner	11.7% (Zeiler/Clarifai)
	Imagenet 2014 winner	6.7% (GoogLeNet)
After ディープ ラーニング	Baidu Arxiv paper: 2015/1/3	6.0%
	Human: Andrej Karpathy	5.1%
	MS Research Arxiv paper: 2015/2/6	4.9%
	Google Arxiv paper: 2015/3/2	4.8%

2015年2月には人間の精度を超えた

画像認識で人間の精度を超えるとは
数年前には考えられなかった

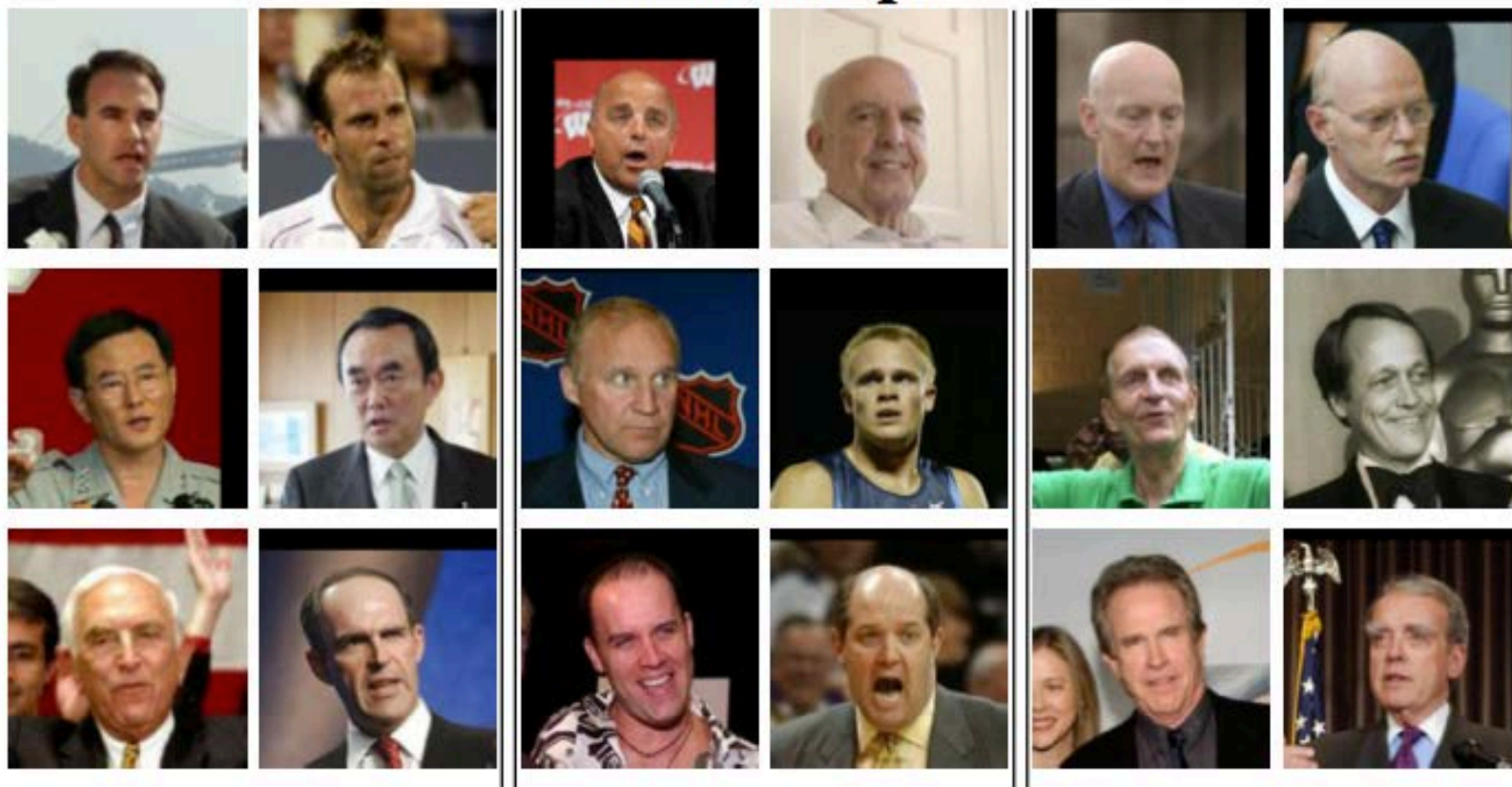
人間を超える画像認識とは？

- Florian Schroffら (Google) の研究。2015年3月
- FaceNet: A Unified Embedding for Face Recognition and Clustering
 - 顔認識と顔画像のクラスタリング
- 800万人の異なる人間の2億枚の顔画像
- 22層のニューラルネットワーク
- 精度: 99.63% ± 0.09 (10分割交差検定)



間違ったケースの全て(別人を同一人物と判定)

False accept



・人間が見ても判定が難しい

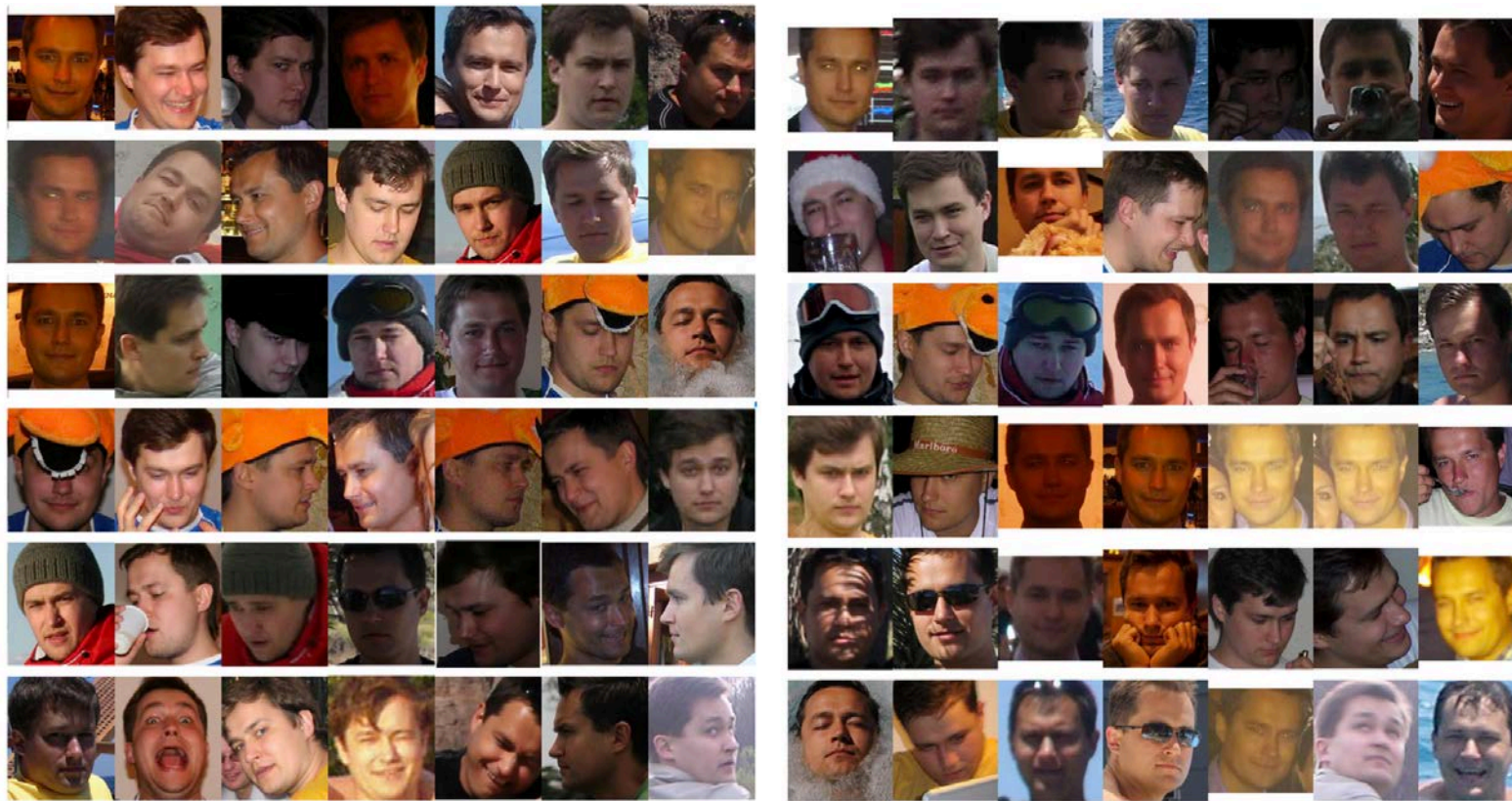
間違ったケースの全て(同一人物を別人と判定)

False reject



・正解データの信ぴょう性のほうが問われる。

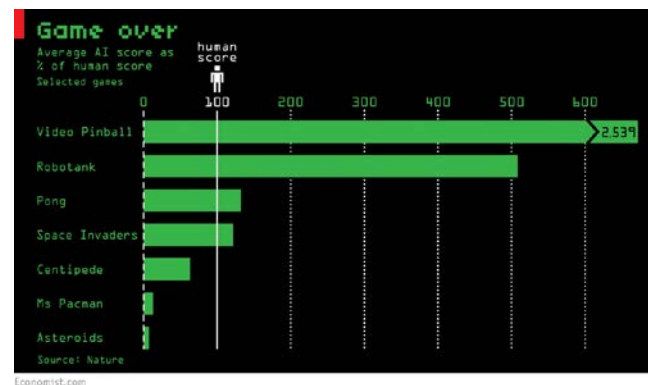
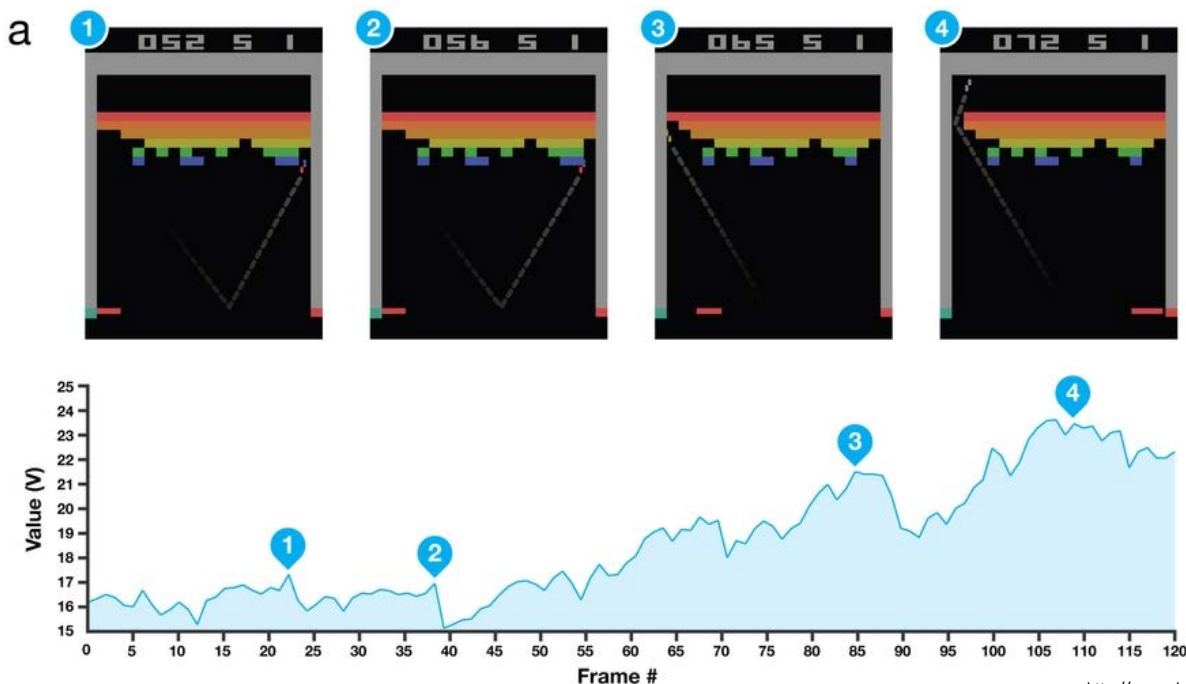
顔画像のクラスタリング



- ・同一人物の写真を正確にグループ化することができる
- ・写真の明るさや角度、表情、サングラスや帽子の有無に対しても頑健

ディープラーニング＋強化学習(2013-)

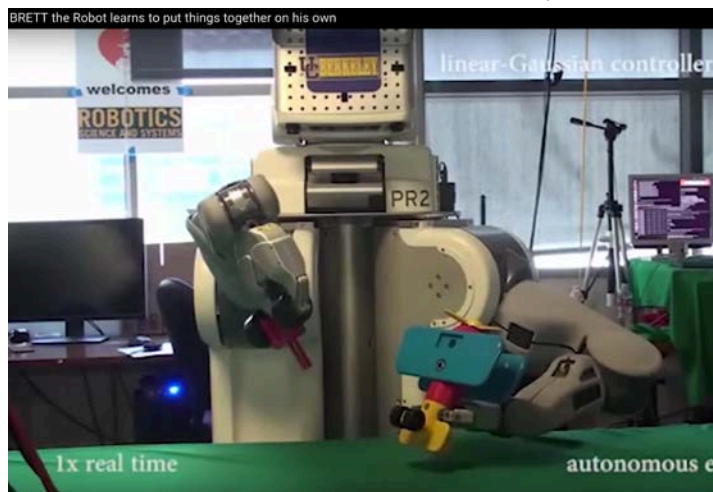
- 強化学習とは、行動を学習する仕組み。
 - 時間遅れで「報酬」が得られるときに、報酬の期待値を最大化する問題として定式化される。
 - 状態と行動の組に対して、望ましさを表す値(Q値)を割り当てる(Q学習)
 - 古くからある技術だが、これまでは、状態の表現を人間が決めるしかなかった
- DeepMindの研究者(D. Hassabisら)は、状態の表現として、ディープラーニングによる画像の特徴量を用いた。
 - スコアを報酬とし、ゲームをプレイするAIとして実現(2013年)→2014年にはGoogleが買収
- 試行錯誤することによって、運動の習熟ができる
 - 最初は下手。繰り返すうちに、うまくなる。
 - 最終的には、ブロック崩しでの通路を作る方法や、インベーダーゲームでの「名古屋撃ち」も身に付ける。
 - 「全く同じプログラム」で、異なるゲームが学習できる
 - 半数以上のゲームで人間のハイスコアを上回る



ディープラーニング＋強化学習：実世界へ(2015-)

- DeepMindのHassabi氏(2015年3月)
 - 「家庭向けロボットは、家の中を掃除したり家事をすることを想定しているが、ここにDeepMindを適用すると自律的な意思決定が可能となる。現行ロボットはプログラム外の事象に遭遇するとうまく対応できないが、DeepMindを搭載すれば自らが驚異的な速度で学習を重ね、幅広い事象への対応が可能となる。」
 - 「現在、自動運転車の開発・テストが世界的に進められているが、ここにDeepMindを適用すると、輸送システム全体が革新的に変化する可能性を秘めている。」
- 実世界への適用
 - 2015年5月 試行錯誤で部品の取付を習熟するロボットの開発(UC Berkeley)
 - 2015年5月 試行錯誤で運転を習熟するミニカーの開発(PFI社, 日本)
 - その他、メリーランド大、EUのプロジェクト等も進展
- 考えてみれば当たり前
 - 犬や猫でもできる。高次の言語能力は必要ない。視覚的な特徴量の抽出が問題だった
 - 歴史的には、多数の人工知能研究者がこのことを主張してきた。表象なき知能、身体性、認知発達ロボティクスなど。知能は環境との相互作用であるという立場。

試行錯誤で作業学ぶロボット(UC Berkeley)



試行錯誤で運転を学習するミニカー(PFI社、日本)



モラベックのパラドックス(1988)

- 長年、人工知能の分野で言われていたこと:「子供のできることほど難しい。」
- ハンス・モラベック、ロドニー・ブルックス、マービン・ミンスキーらが提唱。
- 高度な推論よりも、認識や運動スキルの方が難しい。
 - 1960年～70年代に、定理証明、チェス、医療診断などができている。
 - 一方で、画像認識や「積み木を上手に積む」ような人工知能は一向にできなかった。
- 「コンピュータに知能テストを受けさせたりチェッカーをプレイさせたりするよりも、1歳児レベルの知覚と運動のスキルを与える方が遥かに難しいか、あるいは不可能である」(モラベック)
- ところが、ここ3年くらいの間に
 - 画像認識で人間の精度を上回った。
 - 運動の習熟ができるようになった。

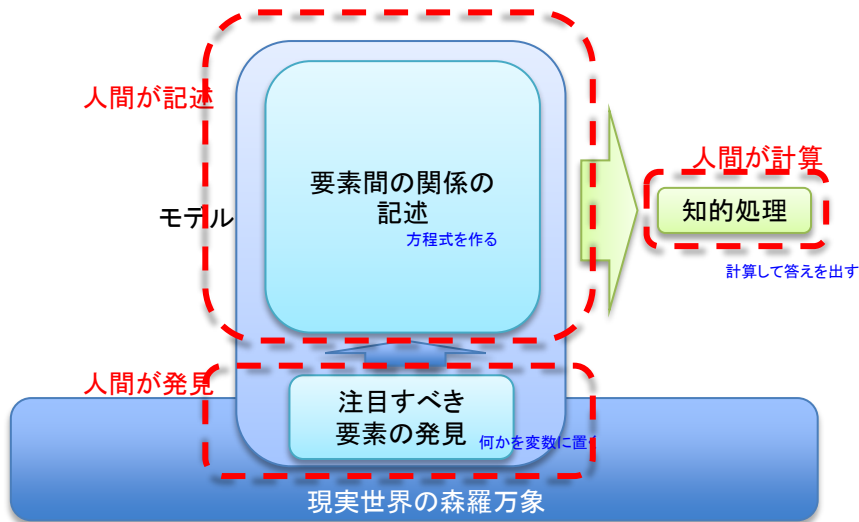
モラベックのパラドックスが崩れた。

現実世界の森羅万象から、特徴量を抽出するところが、実は最も計算量が大きく大変だった。そこができるようになった。

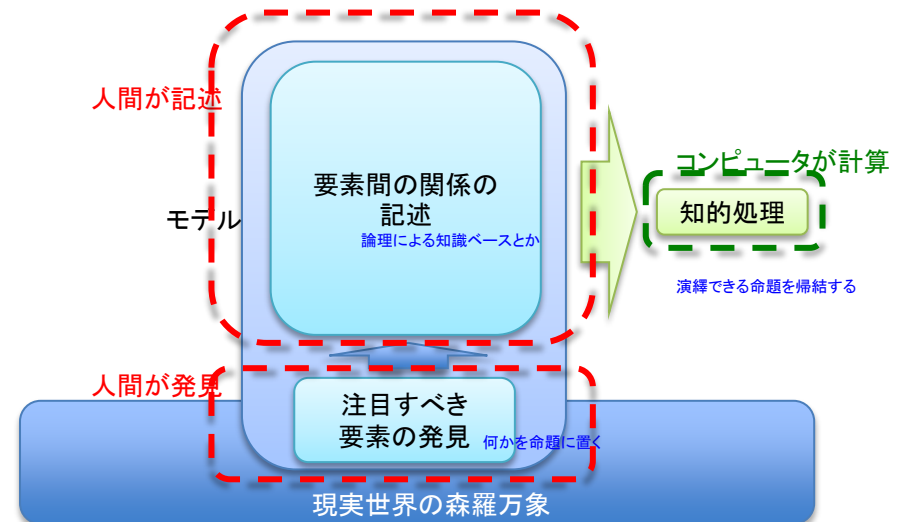
今後、これまでに実現されてきた人工知能の技術と組み合わせられて、急速に進展することが予想される。

→ 後述の「子どもの人工知能」

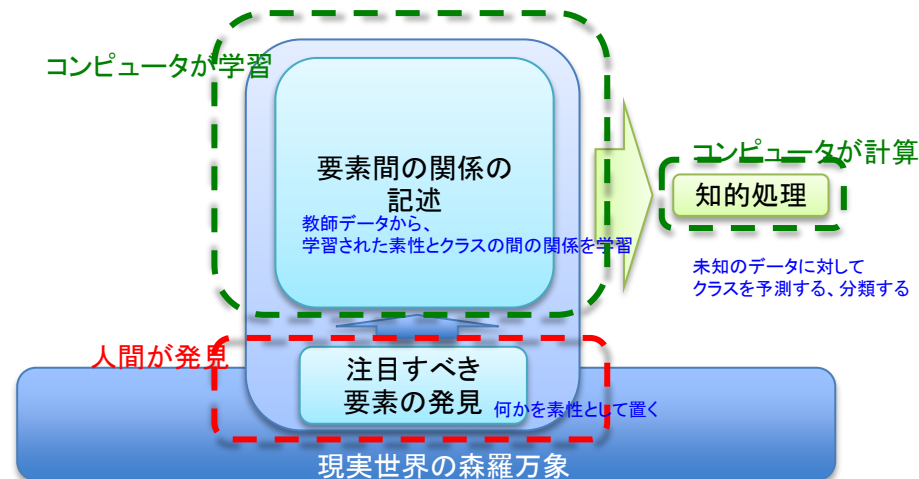
1. 人工知能以前



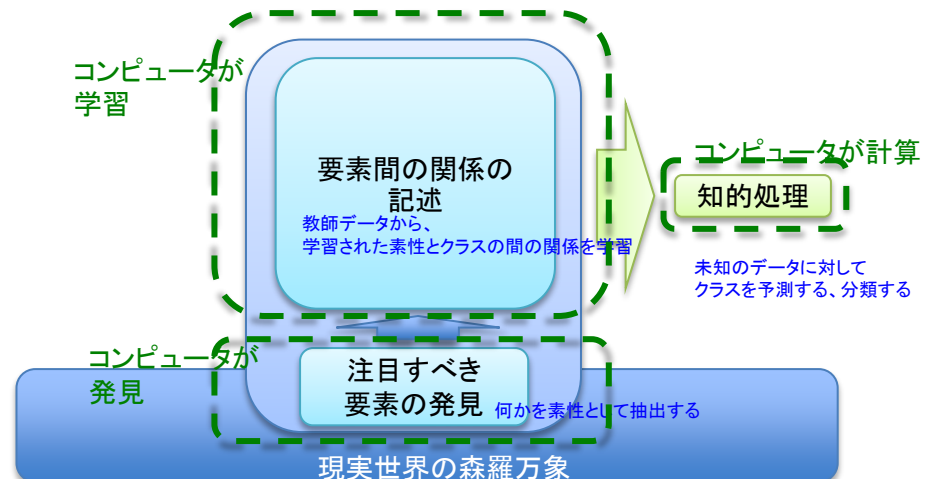
2. 昔の人工知能



3. 機械学習



4. ディープラーニング



「子どもの人工知能」と「大人の人工知能」

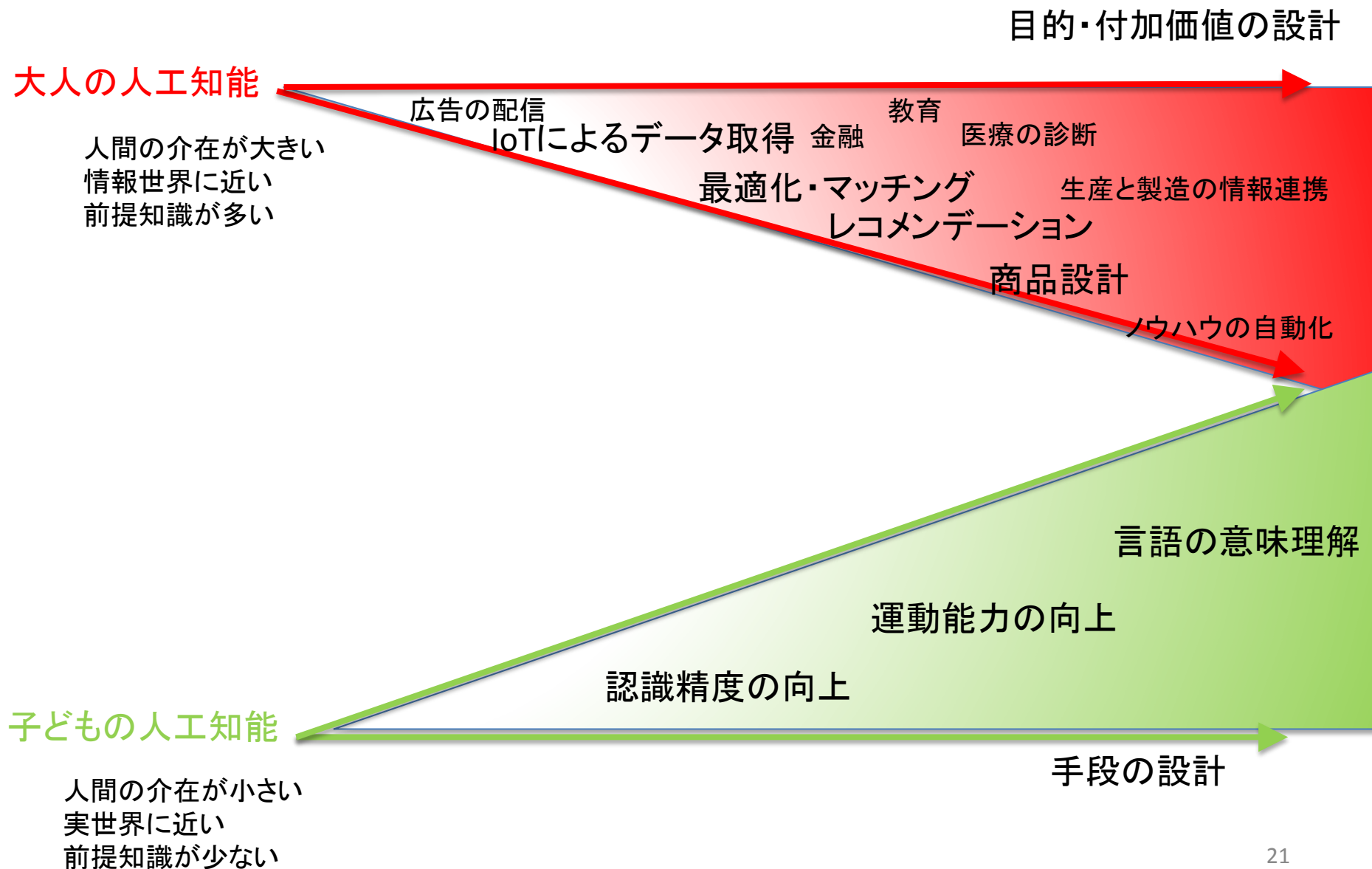
- **大人の人工知能:ビッグデータから人工知能へという持続的イノベーション**
 - ビッグデータ全般、IoT全般、ワトソン、Siri、Pepper...
 - 一見すると専門家(大人)ができることができるが、人間が裏で作りこんでいる。(特徴量の抽出、あるいは関係の記述をしている。)
 - かなりの専門知識、背景知識を最初から組み込んでいる
 - 新たにデータが取れるようになってきた分野では、機械学習の技術が大いに役立つ。
 - 販売、マーケティングなど。今後は、顧客管理、医療、金融、教育など
- **子どもの人工知能:ディープラーニングを突破口とする破壊的イノベーション**
 - ディープラーニングを中心とする発展
 - 子どものことができることができるようになっている
 - 背景知識がほとんどいない状態からの学習
 - 人間の発達と同じような技術進化:認識能力の向上、運動能力の向上、言語の意味理解という順で技術が進展する

特徴量の設計を人間がやらないといけないのが大人の人工知能、
やらなくてよいのが子どもの人工知能

「子どもの人工知能」「大人の人工知能」の違い

- 認識や子どもができることに関して、これまでのように、「研究室の中だけ」「デモの時だけ」「決まったやり方でだけ」できるのではない、本当にできるようになる。
- これまでうまくいかなかった例：
 - 画像認識は、とても人間の認識精度に至らなかった。
 - 音声認識は何度も言い直す必要があった。
 - 翻訳はほとんど使えなかった。
 - 対話はできなかった。
 - ロボットは転んでいた。
 - 物体を認識し、正しく掴んで置くこともできなかった。
- 自動運転を例に
 - 現在の技術は、高精度な自車位置の測定技術＋精密な地図情報
 - 地図がないと走れない
 - 市街地(環境がよく変わる)を走れるのか？郊外は？環境が大きく変化した時は？
 - 大人の人工知能は、例外・環境の変化に弱いのは、歴史的に分かっている
 - 子どもの人工知能のアプローチでは、試行錯誤を通じて運転を学習

長期的に融合する2つのアプローチ



より詳細なマップ

言語知識からの学習

対象物のモデルをベースにした推論の実現

イメージ操作を含んだ空間的理解

関係性の認識

動画の異常検知手法の構築

膨大な動画(VTR等)の索引付け
生成モデルによる補完:
空間のナビゲーションや製品のデザイン

シンボルグラウンディング

DLによる囲碁

DNNで学習したノウハウの言語化、
人間のノウハウのDNN化

認識精度の向上

自動キャプション付け

人間を超える画像認識の精度

各推論方式(演繹や帰納、仮説推論)のDLによる実現

大規模な科学データに対する適用

DL+強化学習+推論

DL+強化学習

シミュレーション精度の大幅向上

人間が構築したオントロジーとの
NNのマッピング技術

情報システムのセキュリティ向上技術の構築

運動能力の向上

膨大な科学論文に対するmemory network版質問応答

日本語でのmemory network版質問応答

多種のデータの異常検知手法の構築

LSTM等による日本語・英語翻訳

多種のデータの演算を含んだ表現学習による分類・回帰の精度向上

数字の演算等のネットワーク

言語の意味理解

数値データの処理

多種のデータのDLによる分類・回帰の精度向上

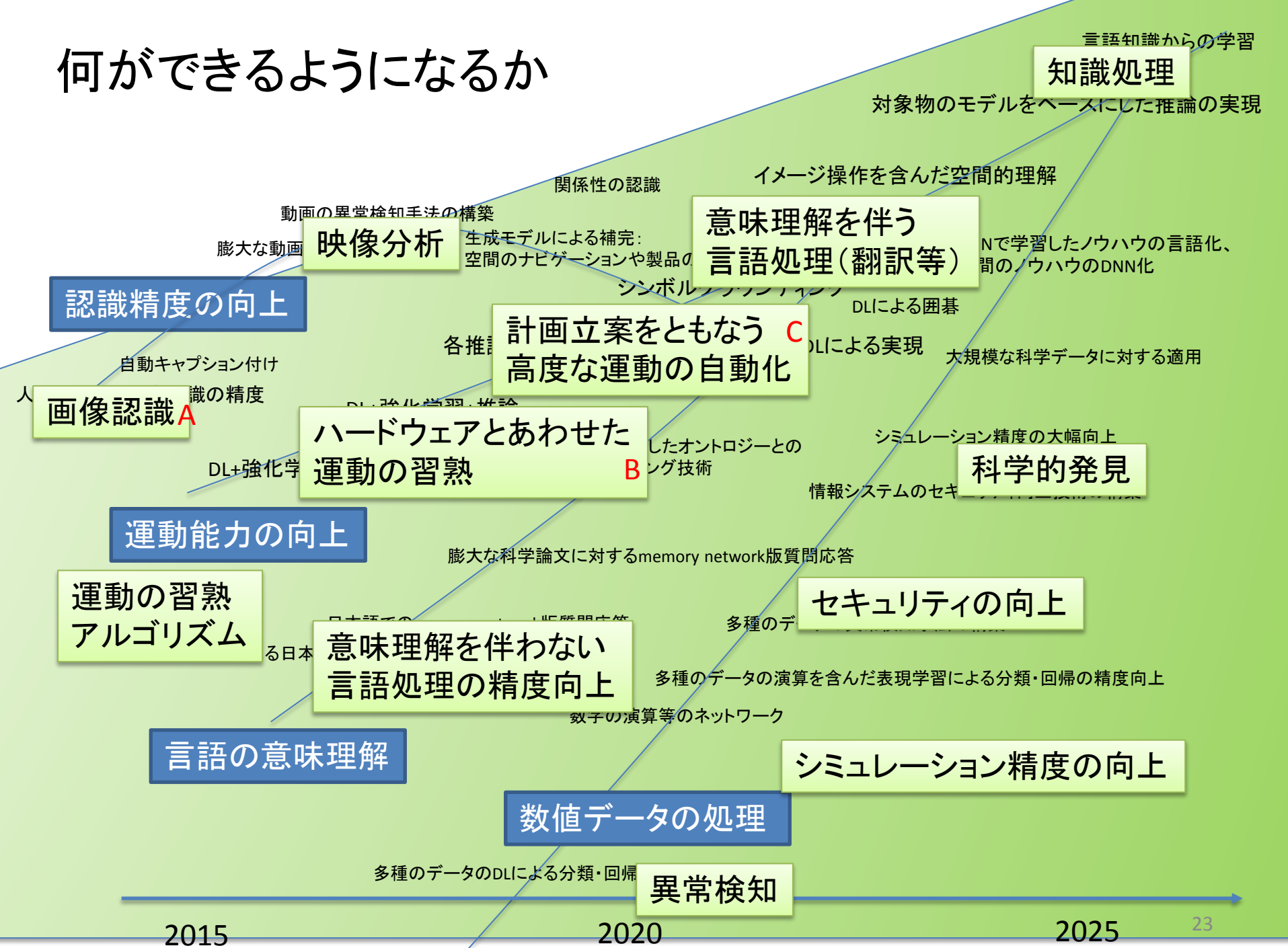
2015

2020

2025

22

何ができるようになるか



どこが日本のチャンスか

- 大人の人工知能
 - 販売、マーケティングと相性が良い。
 - Google、Facebook、Amazon等が強く、日本企業は逆転が難しい
- 子どもの人工知能
 - 認識、運動を伴うので、センサやアクチュエータが必要。
 - 製造業と関連が深く、日本企業にチャンスが大きい
 - 日本は、少子高齢化しており、「運動を伴う労働」のニーズが高い。
 - 農業従事者、建設・物流、介護、廃炉、熟練工の後継者、etc
- 子どもの人工知能のなかでの可能性
 - 画像認識、言語等もあるが、運動能力の向上を主軸にするものが、最もマーケットが大きいのではないか。かつ日本が得意ではないか。
 - 日本人向き：
 - ソフトウェアだけでなくハードウェアとのすり合わせ
 - 性能向上による付加価値向上
 - 自社内のデータで完結する
- したがって、A, B, Cという順序で考えてみる

既存産業の発展

農業

収穫判定

トラクター、コンバインの
適用範囲拡大、効率向上
選別調製等の自動化

自動での収穫
自動での耕うん

建設

測量

掘削、基礎工事、
外装内装作業等の
効率向上

多くの作業の
自動化・効率化

食品
加工

振り分け
確認

カット、皮むき、解体等
の自動化

多くの加工工程の
自動化

組み立て
加工

目視確認の
自動化

動作効率の向上

段取りの自動化
セル生産の自動化

⋮

A: 画像認識

B: 運動の習熟

C: 計画立案を伴う運動

すでにある取り組み

／DLと組み合わせることで飛躍が予想される取り組み

- A: 画像・映像認識
 - － 認証系(顔でログイン、顔パス): マスターカード(米)
 - － 農業(画像による収穫判定): Garuda Robotics(新)、PrecisionHawk(米)
 - － 医療(画像による診断): Enlitic(米)
 - － マーケティング(店舗や街における行動分析): ABEJA(日本)+三越伊勢丹
 - － 画像認識プラットフォーム: Nervana Systems(米)、セーフイー(日)
 - － 交通状況の監視、事故の予防、防犯
 - － ホームセキュリティ、ひとり暮らしの見守り、病院における看護
 - － 災害予兆の検知
- B: 運動の習熟
 - － 産業機械、工作機械の熟練化: Fanuc+PFI(日)、UC Berkeley(米)、(DFKI(独))
 - － 物流の完全自動化:(Fetch Robotics(米), Amazon(米))
 - － 農業:(DFKI(独)、Agriboticx(米))
 - － 介護:(DFKI(独)、サイバーダイン(日))
 - － 家事、調理: ミシガン大(米)
 - － 建設:(コマツ(日)、RedZone Robotics(米))
 - － ロボットの極限環境での作業(原子炉、深海、鉱山、宇宙、災害救助):(DFKI(独))

変化の本質は何か

- 画像・映像認識

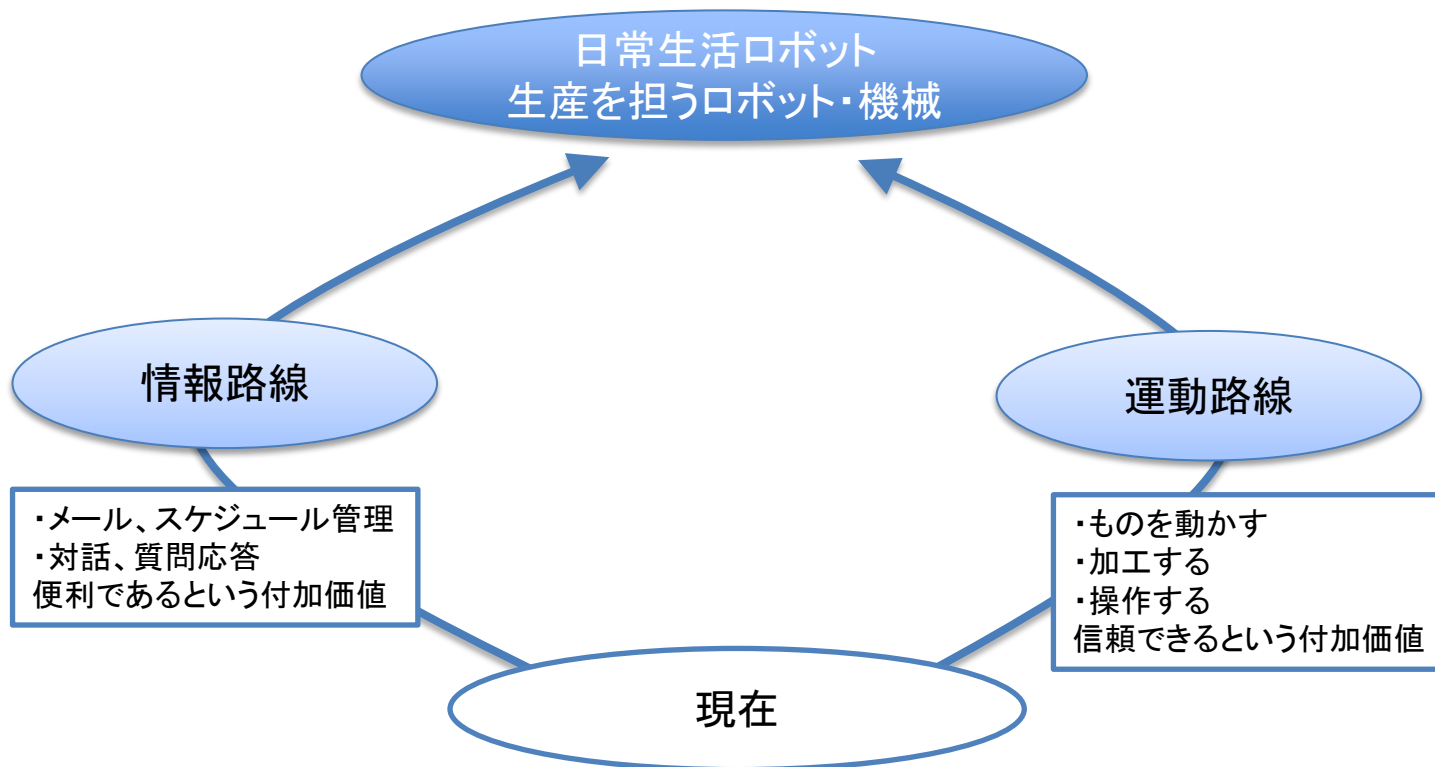
- 世の中に、画像認識ができないから人間がやっている仕事がたくさんある。そこが自動化される。
- コストが下がる。少なく見積もっても、監視のコストは100分の1以下になる。
- それによって、ペイするものが一気に増える。隠れていた潜在的な価値を見る必要

- 運動の習熟

- 我々は、機械は「機械的な動き」しかできない、ロボットは「ロボットの動き」しかできないと思い込んでいる。(まさにこの形容詞が表している。)それが変わる。
- コストは下がるが、設備コスト次第。徐々に価格がさがってくる。
- 既存産業は、部品の時点で規格化され、「機械的な動き」によっても、ある程度の自動化がされている。あるいは、液状化し、成形加工する。今後、自動化できてない部分が自動化できる。
- 自然物を相手にしているものは場面場面で状況が異なるので、そもそも自動化が難しかった。例えば、農業、建設、食品加工。これらは産業としても大きく、ここを勝ち上がった企業が、日常生活・仕事用の機械・ロボットの市場を狙うことができるのではないか。

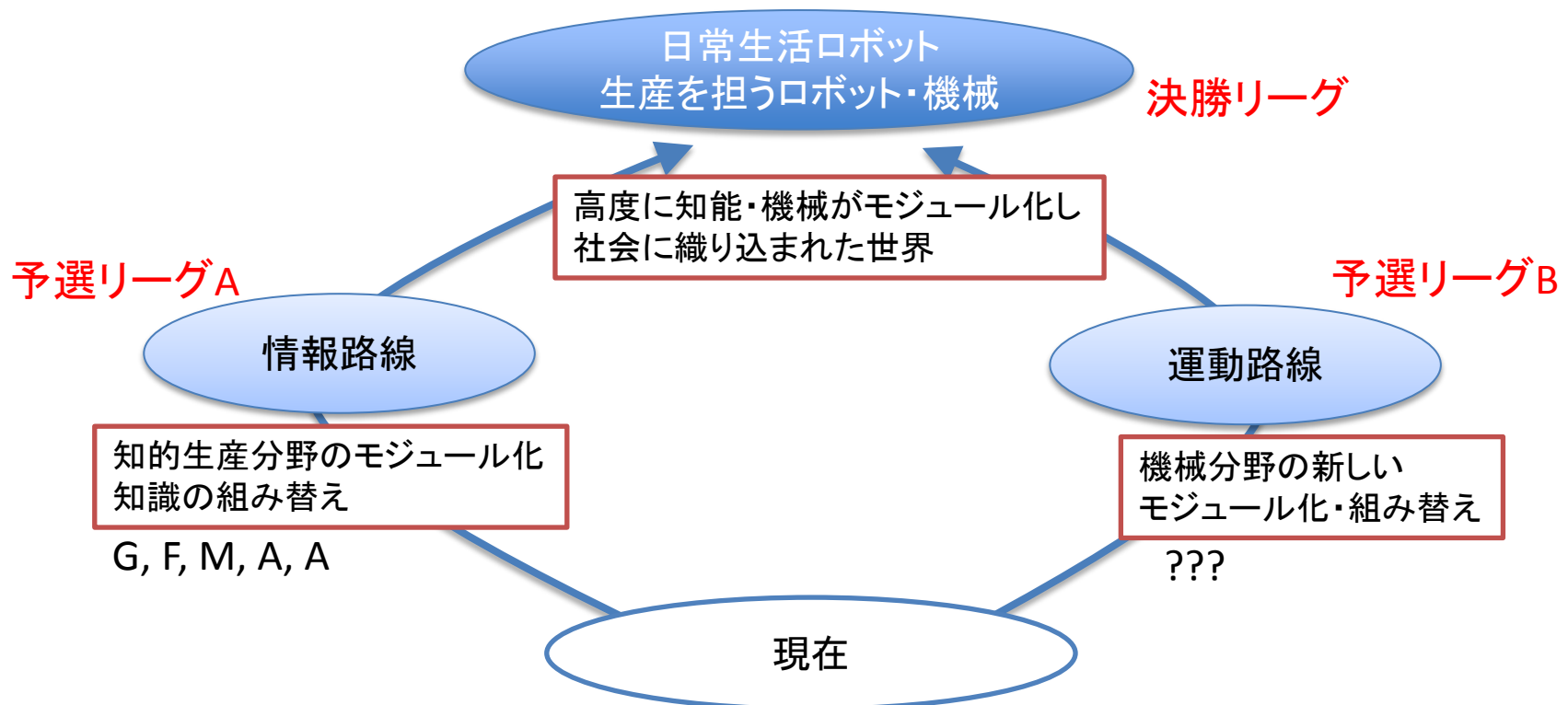
産業としてみたときの方向感

- 最終的には、日常生活、仕事におけるロボット・機械の活用。状況ごとに個別性があるので、特徴量生成の能力がない状況では対応できなかった。この自動化にどう至るかが鍵。
- 情報路線で行く道（Google, Facebook系）と、運動路線で行く道があるのではないか。
- 海外企業・研究者は機械系・ロボットに苦手意識：cf) 2015 ICML deep learning workshop パネル



産業プラットフォームの変化

- 運動路線系のモジュール化が進むはず。
 - これらが効きやすい産業: 食品加工、農業、建設。ハードウェア(素材、駆動系)も重要になる。
- 個別領域でどうデータを獲得するかが鍵
 - アルゴリズム自体は競争力の差異にならない。データの取得が参入障壁を生む。
 - データを取るために、画像認識の製品投入→機械・ロボティクスの投入が有望な戦略では
- 一方で、情報路線系のモジュール化も進む
 - ここは、英語圏のマーケットを抑えている企業に勝つのは難しいのでは
- 予選を勝ち進んだ企業が決勝に進むイメージ



今後やるべきこと

- 人工知能の研究開発投資
 - 特に子どもの人工知能への研究開発
 - 早期に試作品を作り、その可能性を世の中に見せる
 - 産業化だけでなく、知能の本質に迫る基礎研究も重要
 - 特に、表現の学習とは何かというのは、科学の本質に迫る深遠な問題
- 新しいチャンスを捉える
 - ベンチャー育成
 - 既存の企業内(人工知能と全く関係ないと思っているような産業)でも大いにチャンスがある。事業がどう変わるかを見抜き、先行投資
- 人材を創る
 - 高度な人工知能(ディープラーニング等)を使いこなす人材の育成が急務
 - 企業の技術者の研修も重要
- 未来の像を創る
 - 人間の幸せを実現する、明るい未来社会を描く。
 - 技術の進展に正しく基づいた未来像の発信:コンテンツの力

まとめ：人工知能による「ものづくり」の復権へ

- 少子高齢化しており、「運動を伴う労働」のニーズが高い。
 - 農業従事者、建設・物流、介護、廃炉、熟練工の後継者、etc
- 「子どもの人工知能」は、広い意味でものづくりと相性がよい
 - 日本の強みを活かせる
- 人工知能は、良い条件が揃っている
 - 人工知能研究者の人数が多い：知能の本質を考え続けた人が多い。
 - 世代を通じた理解：大きな組織も動きやすい
 - 言語がハンデにならない：アルゴリズムを製品とすり合わせる世界
- チャンスを捉えるには、正しく動いていくことが重要
 - この時点における産業政策は極めて重要
 - 社会全体で新しい社会の未来像を描いていくことが大事