

4つのAI活用

- **既存業務の効率化や高度化**
今まで人が対応していた業務をAIで置き換えたり、人をAIでアシストする
- **AIによる新規サービスの立ち上げ**
AIでデータ分析を行い、その結果を利用して新サービスを創造する
冷蔵庫メーカーが生命保険会社と競合する！？
- **ビジネスのスケールアップ加速**
AIによりビジネスをスケールアップさせる
- **リアル世界のインテリジェント化**
AIを適用すれば根本から改善できるサービスは何か？
東大 松尾先生・・・AIの深化によりモノやサービスが目を持つようになる
(交通状況に合わせて、信号をダイナミックに切り替える、など)

誰もが高度に知識にアクセスする時代
知識からスピードの時代へ・・・AIで加速 2位ではダメ
AIによる自動化 と ネットワーク効果
正確で迅速な判断・・・予測による先読み

AMAZONではユーザーが商品を購入する前に、
出荷を行うサービスを準備中

都市機能AIで高度化

NTT 防犯や防災、まず米で

【ラスベガス＝堀越功】NTTは人工知能(AI)を含むシステムを導入し都市の機能を高める取り組みを全世界で展開する。第1弾として米国ラスベガス市と連携し、犯罪や災害を防ぐ取り組みを検証する。NTTは海外事業の強化を今後の成長戦略の柱として掲げている。ラスベガス市との取り組みの成否が海外戦略の将来性を占う一つの試金石となりそうだ。

「ラスベガス市はスマートシティ化を推進する。公共安全分野のほか、教育など幅広い分野で連携を進めた」と沢田副社長は語った。NTTがシステムを導入し都市機能を高める「都市のスマート化」で国外の都市と連携するのは初めてという。

「米市民の間で公共安全の要求が高まっている。新たな収益の流れをつくりたい」。NTTとの連携を決めたラスベガス市

のIT担当ディレクターであるマイケル・シャールウッド氏は話す。ラスベガス中心部で17年に銃乱射事件が起き、多数の被害者が出たことは記憶に新しい。住民を守り、都市のスマート化によって経済を活性化する点も含めて両者の連携は視野に入っている。シャールウッド氏は「NTTは我々が望むことを実現できる連携相手であり、広い分野で深く協業を進めたい」と続ける。

9月からラスベガス市で実証実験を開始する。NTTグループ傘下の南アフリカのIT大手ディメンション・データやNTTデータなどのほか、米IT大手のデルテクノロジーも協力する。

土木分野でのAIの適用例

特集 維持・補修2017

人工知能 AI

助けて 点検・診断の自動化は 夢物語じゃない!

大水深や視界不良といった難条件でも施工できる水中ロボット (メーカー)

画像を基に舗装の不具合を抽出 (建設コンサルタント)

与条件を踏まえた最適な橋梁形式の選定 (建設コンサルタント)

図面が存在しない既設構造物の仕様を推定できないか (建設)

シールドトンネル工事の自動化。ルートを入力すればマシンを自動運転する。セグメントの選定・組み立ても自動化できればいい (建設)

設備の自動運転 (建設)

デジタルカメラなどで撮影した画像から、コンクリート構造物のひび割れを精度良く検出したい。現在のアルゴリズムは実用レベルと言いがたく、自動検出に対するニーズが高まっている (点検)

トンネル工事における地山の前探査に活用 (建設)

設計照査を効率化したい。発生パターンをAIに学習させて、「よくあるミス」の発生を防止する (建設コンサルタント)

難易度が低い構造物の設計はAIで自動化したい。難易度が高い構造物は、熟練技術者によるオーダーメイドがよい (建設コンサルタント)

労務管理に活用したい。協力企業や作業班、作業員ごとの特性、出来高、歩掛かり、賃金、健康状態、傷病歴を統合管理 (建設)

道路交通量調査を効率化したい (建設コンサルタント)

音声認識による図面作成、プレゼン資料の作成 (建設)

事故事例を収集しデータベース化。現場で対話型AIに作業内容を伝え、適切な安全対策を回答するようできないか (建設)

鉄道周辺にある「当社以外」の用地について、変状箇所を自動的に調査し、相手に通知できるようにしたい (鉄道)

災害発生危険度の推定 (建設コンサルタント)

道路パトロールのルート最適化や、除雪の最適化(出動者、タイミング、ルート) (建設コンサルタント)

図面上から数量を算出 (建設)

施工時に、様々な計測データから異常の発生を予測 (建設)

ボーリングは限られたポイントでしか実施できない。過去の地歴や隣接するボーリングデータなどから、施工箇所の地質データをより正確に判別したい (建設)

人が土木構造物に近づかなくても、遠隔で自動的に修繕できるようにしたい (鉄道)

工事用車両の運行を最適化したい (建設)

点検結果やセンサーで取得したデータなどから、構造物の余寿命や耐荷力などを推定する (建設コンサルタント)

ドローンで撮影した画像から地滑りや土砂崩れが発生しやすい箇所を抽出し、法面の防災対策を提案 (専門工事)

工程や人員配置、機械設備の配置、調達時期・ルートなど、施工計画の最適化 (建設)

労働災害の防止に活用しては、職人の練度や工程、当日の朝夕の表情データを用いて危険を予知 (建設コンサルタント)

将来の土地利用状況の予測 (建設コンサルタント)

現場で撮影したデジタル写真を自動的に整理してくれると助かる (建設)

地形や水位、河川断面、降雨の傾向から河川氾濫の発生を予測 (建設コンサルタント)

人手や予算の不足に悩む維持管理の現場にAI(人工知能)がやってきた! ひび割れを自動で検出したり、劣化を予測したりと用途は様々。積年の課題を解決すべく、様々な企業が開発にまい進している。AIの導入に関心があるあなたも、実力に懐疑的なあなたも、まずは本特集で様々な事例に触れ、その可能性を感じ取ってほしい。(木村 駿)

掘削排土の映像からN値を推定したり、地層を判別させたりして、自動的に土質柱状図を作成 (専門工事)

設計照査を効率化したい。発生パターンをAIに学習させて、「よくあるミス」の発生を防止する (建設コンサルタント)

トンネル工事における地山の前探査に活用 (建設)

設計照査を効率化したい。発生パターンをAIに学習させて、「よくあるミス」の発生を防止する (建設コンサルタント)

難易度が低い構造物の設計はAIで自動化したい。難易度が高い構造物は、熟練技術者によるオーダーメイドがよい (建設コンサルタント)

事故事例を収集しデータベース化。現場で対話型AIに作業内容を伝え、適切な安全対策を回答するようできないか (建設)

道路パトロールのルート最適化や、除雪の最適化(出動者、タイミング、ルート) (建設コンサルタント)

図面上から数量を算出 (建設)

ボーリングは限られたポイントでしか実施できない。過去の地歴や隣接するボーリングデータなどから、施工箇所の地質データをより正確に判別したい (建設)

工事用車両の運行を最適化したい (建設)

点検結果やセンサーで取得したデータなどから、構造物の余寿命や耐荷力などを推定する (建設コンサルタント)

図面を分析し、施工時にヒューマンエラーが発生しやすい箇所を特定。対策を講じる (建設)

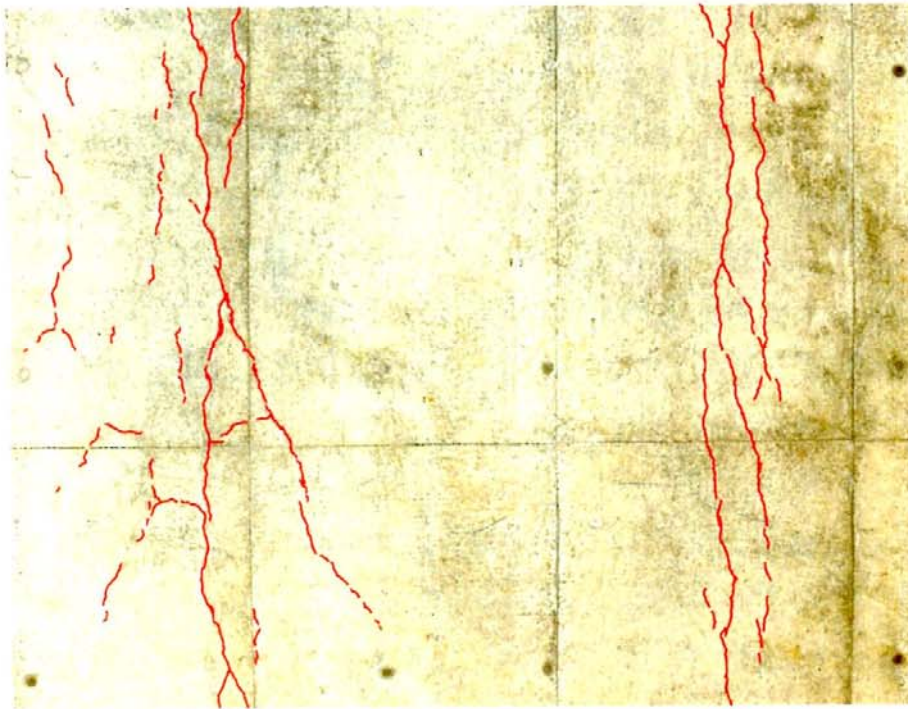
CONTENTS

- PART1 活用動向 維持管理とAIは相性相愛!?▶38
- PART2 Q&A 5分で分かるAIのキホン▶44
- PART3 事例総覧 秒読み開始! AIの現場導入▶48
- PART4 開発方法 誰がどう開発する?▶56
- PART5 先端研究 「瞬間診断」の時代が来る▶58

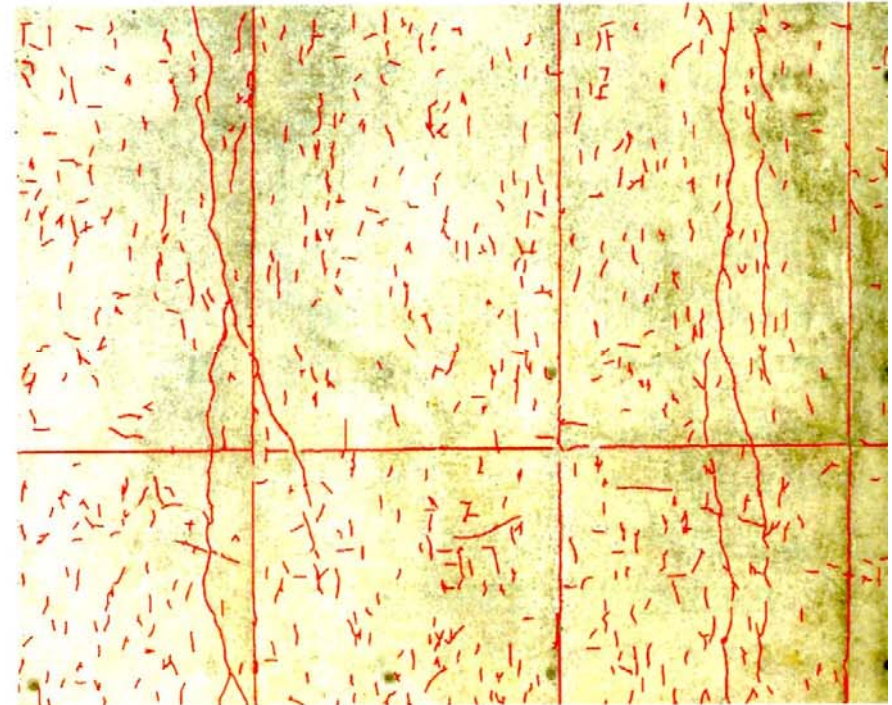
■ 調査 ■ 計画・設計 ■ 施工 ■ 維持管理

本誌が約50人の土木技術者にAIで解決したい課題を調査した結果を、調査、計画・設計、施工、維持管理に分類し、36〜57ページの吹き出し内に抜粋して示した。カッコ内は回答者が所属する企業の業種

道路や橋のひび割れを、人にかわって人工知能が点検



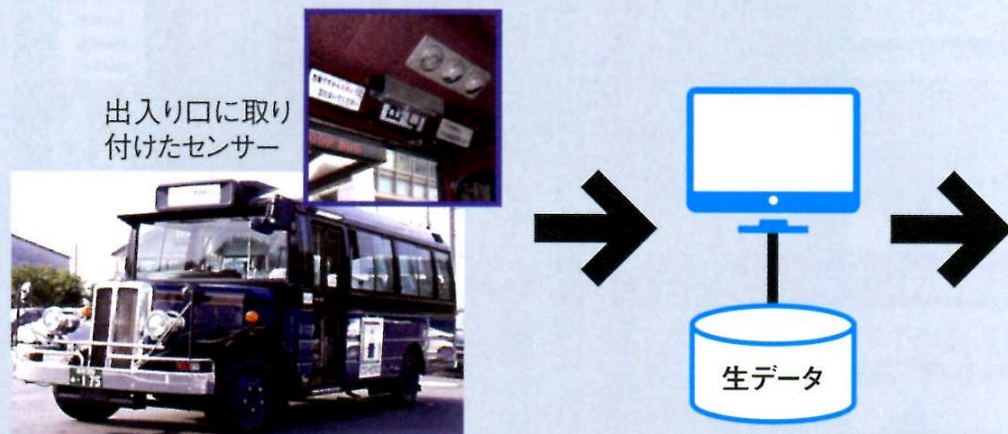
人工知能を使って検出した結果
正解率 80%程度



従来法による検出した結果
正解率 12%程度

産業技術総合研究所／東北大学／首都高速技術 の共同研究
人工知能によるひび割れ検出Webサービスを無料で公開

図2 バスにセンサーを取り付けることで実態が把握できるように
(埼玉県川越市のイーグルバスの事例)



区間ごとの乗客の有無や遅延などが一目で分かるように

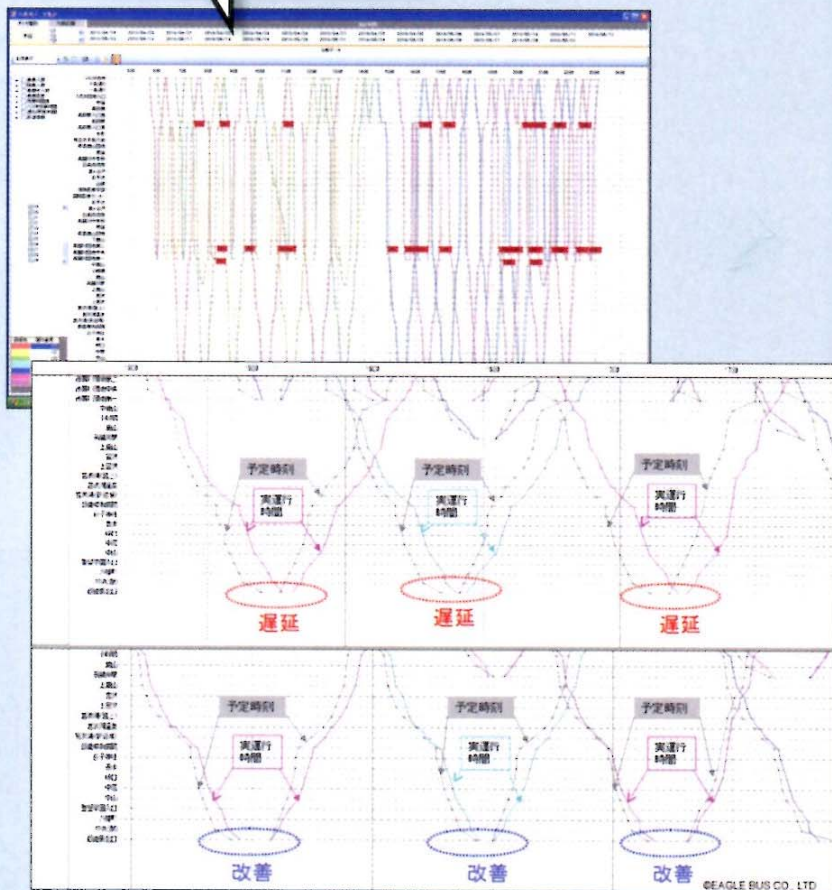


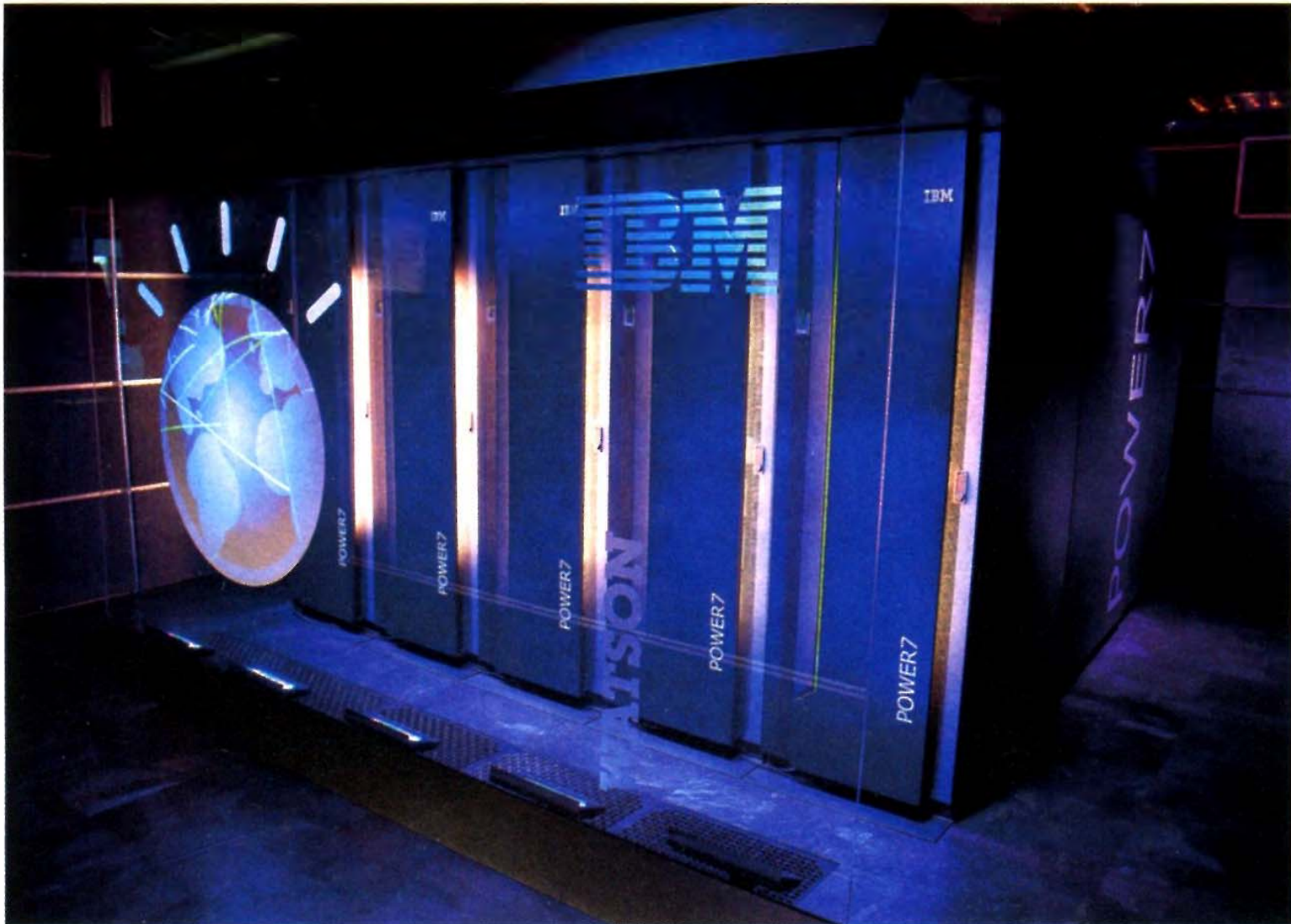
図3 IoTによって各分野で事業の再定義が始まる

- サーモスタット → **スマートホーム**
- 電子商取引 → **物流**
- 〇〇販売 → **〇〇シェア(自動車など)**
- タイヤ販売 → **Pay per Mile**
- 照明機器 → **照明制御/保守**
- 空調機器 → **快適さ/涼しさ提供**

JR東日本も電車の自動運転に向けて、車両、架線、線路、周辺地図などのデータを取得開始

「名探偵」ワトソンの使い道は無限大

独立したデータ同士を分析し、意味のあるものに変換するIBMのワトソン。莫大なデータからワトソン自身が類推し、人間では思い付かないような新しい回答を導き出す。



Newsweek日本版 2016.5.17 「人工知能は人間を超えたか」

ロボットやAIが
もたらすインパクトは大きい

14%

世界32カ国を分析した結果、高度な自動化が可能で、自動化によって消える可能性が高い仕事の割合は14%だった。32カ国でその対象は6600万人に及ぶ

(出所)OECD(2018年)

約1600兆円

2030年までにAIが世界経済に最大15兆ドル(約1600兆円)の経済効果をもたらすと試算されている

(出所)PwC analysis(2017年)

32%

上記の14%を除いた仕事以外でも、32%の仕事に就く人々は、自動化によって仕事の内容が大きく変化するため、労働者の再教育が必要となる

(出所)OECD(2018年)

49%

日本の労働人口の49%が就いている職業は、10~20年後の社会でAIやロボットなどにより代替することが、技術的には可能となる

(出所)NRIとオックスフォード大学の共同研究(2015年)

3億7500万人

ロボットの利用や自動化によって、2030年までに世界で最大3億7500万人(世界の労働人口の最大14%)が異なる業種への転職を迫られる

(出所)マッキンゼー・グローバル研究所(2017年)

人間に求められる仕事が変わる

将来にわたって人間が担う仕事

創造性が求められる	ミュージシャン アートディレクター ゲームクリエイター 映画監督 コピーライター 料理研究家	医療・福祉・介護	ケアマネジャー 外科医 獣医師 精神科医 産婦人科医 福祉施設介護職員 保育士
非定型な業務	マーケティングリサーチャー 経営コンサルタント 商品開発部員 美容師 アナウンサー	教育	学校カウンセラー 教員 人類学者 社会学研究者 学芸員 日本語教師
人間性や協調性が求められる	国際協力専門家 スポーツインストラクター ツアーコンダクター		

AIやロボットへの代替可能性が高い仕事

工業・製造業	石油精製オペレーター 電子部品製造工 自動車組立工 IC(集積回路)生産オペレーター 検収・検品係員	バックオフィス	一般事務員 医療事務員 行政事務員 (国・県市町村) 経理事務員 保険事務員 データ入力係
営業	清涼飲料のルートセールス員		
物流・人流	出荷・発送係員 新聞配達員 宅配便配達員 タクシー運転者 路線バス運転者 倉庫作業員	サービス業	ホテル客室係 レジ係 寮・マンションの管理人 警備員 受付係

(注)それぞれ100種の職業から主なものを抜粋。順不同 (出所)「日本の労働人口の49%が人工知能やロボット等で代替可能に」(NRI, 2015年)を基に、本誌が抜粋・分類

人手不足の心配無用? 早くも来た人余り時代

「銀行の新卒採用は正気のものとは思えない」

金融庁の幹部が不満げに愚痴っていました。「役所が指導すればいいのでは」と突っ込むのも忘れ、大いに賛同してしまいました。メガバンクは新卒の最大の受け皿になっています。今年も3メガがそれぞれ1000人を軽く超える新入社員を迎えました。これだけの人にこれからも仕事を与え続けられるのか、と幹部は心配しているのです。


煎じ詰めれば、信用に基づいて情報をやり取りするのが銀行の主な業務です。これはIT(情報技術)の得意分野でもあります。銀行で人がやるべき仕事はこれからどんどん減っていくはずです。金融とITを融合させるフィンテックの採用を銀行幹部が本気で考えていないのか、新卒の将来を慮っていないのか、いずれにしても心配です。

IoT, AIが
進むと人余り

日経ビジネス 2017年7月

単純な
仕事
が
なくなる

情報化社会の行き着く先としてかねて懸念されていたことが今現実になるうとしています。不景気のせいではなく、生産性が高まったことで余った労働力をどう使い、どのように生産性向上の果実を分け合うか。社会はその時代に備える必要があります。働く時間が減り報酬が増えることになれば、今号の特集「夜」も少しは活性化するのでしょうか。

(東昌樹) 

AIの適用例 製造業 工場、倉庫、運用が変わる

ドイツのインダストリー 4.0 が発端

あらゆる種類の製造機器をICTを活用して、最適化して生産性を大きく向上
世界中で同様の取り組み

アメリカ GE中でIIC (Industrial Internet Consortium)

日本 機械学会「つながる工場」のIVI (Industrial Value-chain Initiative)

中国 中国製造2025 **中国の躍進**

日本はロボット出荷台数で中国に抜かれ、稼働台数でも2016年に抜かれる

ドイツ企業は、Industry 4.0を中国に導入したい

製造業における重要事項・・・少量多品種生産(顧客ニーズ)・・・AIによる生産計画

センサーやAIを活用することで従来とは異なるコンセプトを持つロボットの登場

(AIを活用した生産支援協働型ロボット バクスター・・・ルンバ掃除機の子会社)

日本もファナックが追随している

ドイツのインダストリー 4.0の目標は、製造業をサービス業化すること

モノを販売して終わるビジネスモデルから脱却、製品に付加価値を付けて売上げを向上

定期点検・・・CBM (Condition Based Maintenance) 対象物の状況に合わせて維持管理

機械設備や車にセンサーを付けて使用状況、壊れ方など確認し、補修を行う AI活用

工場内だけでなく、倉庫においてもAIの導入が進んでいる => 物流現場の革新

クレーン方式からロボットピックアップシステムへ・・・照明・空調不要

=> さらに医療、農業、セキュリティ分野へのAI活用に進んでいく

インダストリー 4.0 ドイツ・欧米型

抜きんでた人依存・・・エリートチームが企画や戦略を練り、ブルーカラーが決められた役割を忠実にこなすという動態
生産性は上がるが、格差が生じ易い

日本型？

大部屋でみんなが顔を合わせ、知恵を出し合って企画をまとめるという文化
現場の一人ひとりが考え、それぞれの力を結集・結束させることが強み
日本の現場には、個々の人がこだわりを持って突き詰めようとする傾向がある。
これは過剰な創意工夫となり生産性が下がりやすいという負の側面があるが、優れた技を持った”匠”を生み出す土壌となってきた。

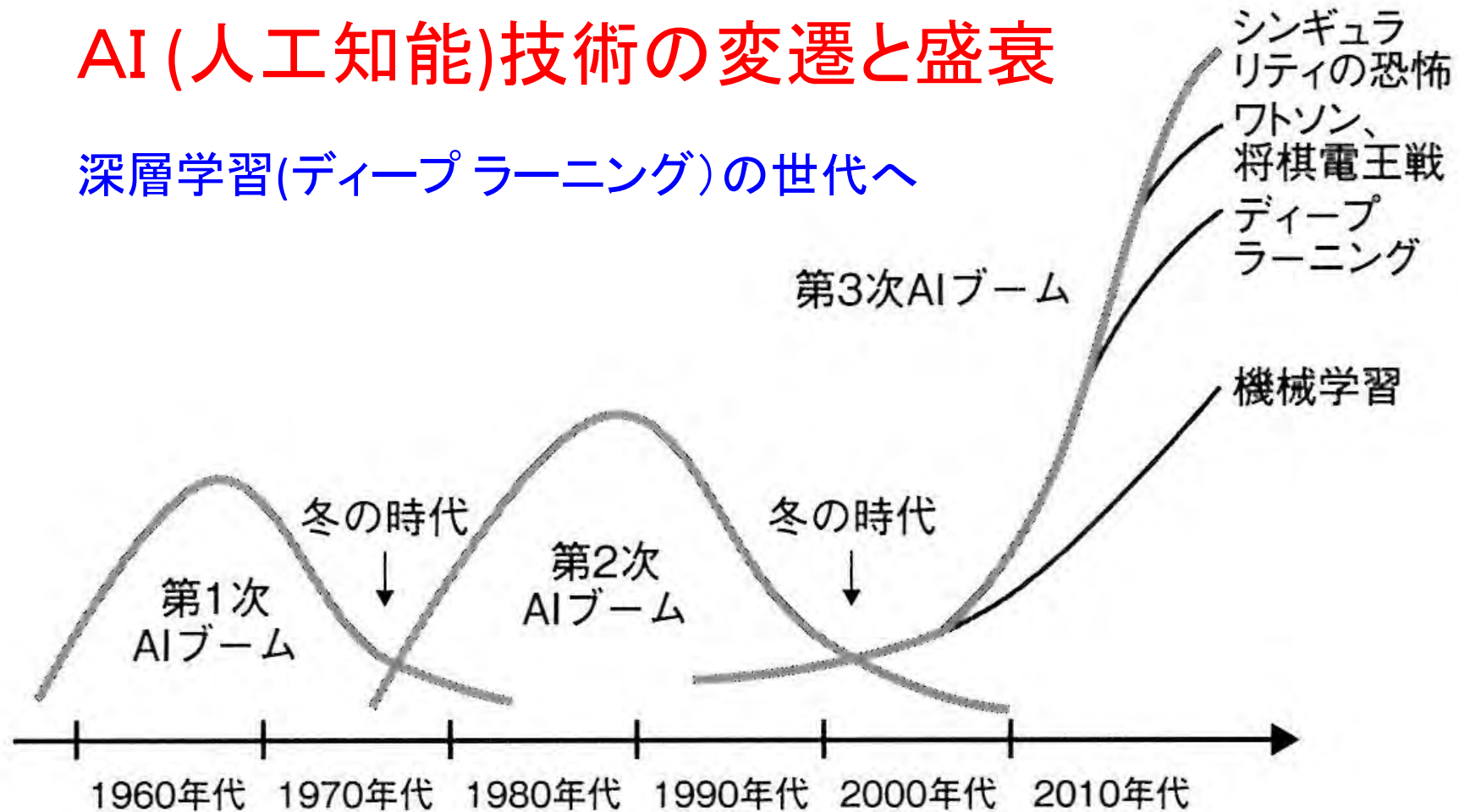
”匠”の存在 ——> AI化

?? AIに任せきりになると、”匠”を生み出すメカニズムが消える
進化しなくなる？

一方、日本企業では、足元の忙しさに忙殺されて、**自分が担当する仕事しか見えていない、前後左右のことを考えられない人が増えている。**対処するには、IoT, AIといった技術を活用し、学習速度を上げ、視野を広げる

AI (人工知能)技術の変遷と盛衰

深層学習(ディープラーニング)の世代へ



松尾 豊:人工知能は人間を超えるか ディープラーニングの先にあるもの (角川EPUB選書)

AI手法の時代変遷

第一次AIブーム

ルールベースのAI
論理的

第二次AIブーム

エキスパートシステム・・・目的に応じた専用マシン
ニューラルネットワーク・・・汎用・精度が上がらない
データマイニング・・・特徴抽出・クラスター分け

第三次AIブームー現在 ー> そろそろピークか？

確率・統計的手法(ベイズの定理を用いた推論)

大量のデータが必要ー>ビッグデータ

Google検索(検索の窓口・・・頻度の多いもの)、Amazon検索

機械学習(自律的推論ーニューラルネットワーク)

マシン・ラーニング, 多層NN Deep Learning

ビッグデーター>ニューラルネットワークのチューニング

Siri(Apple)、Symantec検索(Google), Amazon(グラフ)検索など

(次世代の入カインターフェイス 音声 主導権争い)

Breakthrough
(Deep Learning)



人工知能

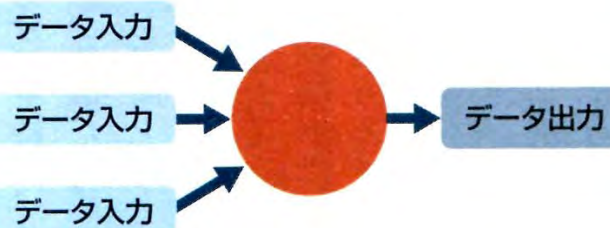
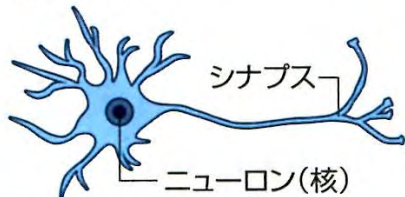
- ニューラル・ネットワーク
 - 教師あり
 - パーセプトロン
 - バックプロパゲーション
 - アソシアトロン
 - 教師なし
 - 自己組織化マップ(SOM)
 - ボルツマン・マシン
- データ・マイニング
 - 主成分分析
 - クラスタ分析
 - サポート・ベクタ・マシン(SVM)
- 強化学習
 - Qラーニング

最近話題の
ディープ・ラーニング
(ディープ・ニューラル・
ネットワーク)も仲間

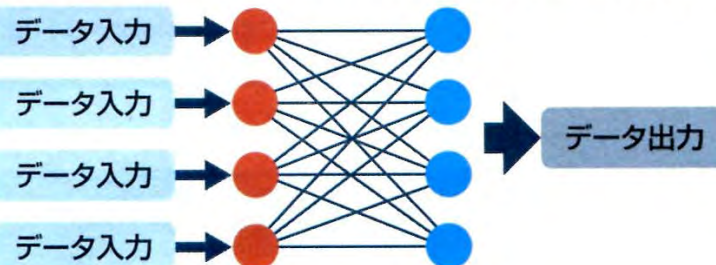
基本2 ディープラーニングは何がすごいのか？

STEP1 脳をまねて作られた形式ニューロン

人間の脳はニューロン(神経細胞)で構成されているから…

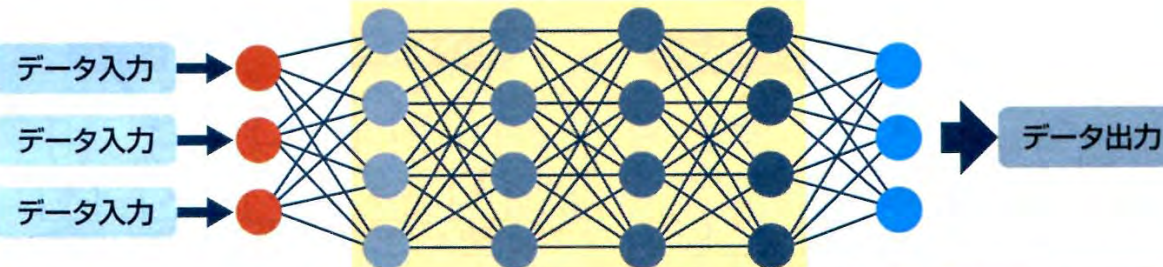


STEP2 形式ニューロンを複層構造にしたパーセプトロン



- スпамメールの分類など、初歩的な機械学習が可能に
- しかし二者択一のできない複雑な問題は解けず、活用できる場面は限定的だった

STEP3 階層の数を増すことで考える力を得たディープラーニング



- 大量のデータと、それを処理できる高性能な半導体・コンピュータの存在で可能になった
- 大量の画像から共通した特徴を抽出するなど、あいまいな問題も解けるようになった

(出所) 取材を基に本誌作成

機械学習、
ニューラルネットワーク、
ディープ・ラーニング
(深層学習)の

基礎知識

機械学習が注目される理由

2012年に機械学習の一種であるディープラーニング(深層学習)というアプローチを用いて、大幅な画像認識の精度の改善が達成された。2016年にGoogle傘下のDeepMindの開発した「AlphaGo」が囲碁の世界チャンピオンを破った。・・・画像認識の精度向上

アルゴリズムの進化	<ul style="list-style-type: none">・ディープラーニングや強化学習など、応用可能性と実行力の高い機械学習の手法の進化が続いているため
データ量の増大	<ul style="list-style-type: none">・インターネットの発展により、帯域の増強と共に画像、映像、音声、テキストなどさまざまなデータが増大しているため・企業における業務のシステム化やセンサーの普及により、さまざまな種類のデータ量が増大しているため
計算資源の進化	<ul style="list-style-type: none">・コンピューターの処理能力が格段に高まっているため（スパコンの処理能力の向上、GPU、TPUなどの開発・普及）
アルゴリズム、データ、計算資源の利用可能性の向上	<ul style="list-style-type: none">・オープンソースのライブラリやTensorFlowなどのツールにより、機械学習やディープラーニングのアルゴリズムが簡単に使えるようになっているため・各種のデータソースが整備され、学習のためのデータが用意しやすくなっているため・クラウドサービスの普及により、高性能の計算資源が安価に使えるようになっているため

機械学習とは何か?

簡単にいえば、「データからパターンやルールを機械自身に見つけさせる仕組み」

”100%の精度を達成することはほとんど出来ない”

データから観測できない情報や正解データ自体の曖昧さ・ばらつきなどにより機械学習の識別や予測の結果と実際の結果はある程度のずれを持つ。・・・「誤差」

機械学習とルールベースの比較

	機械学習	ルールベース
概要	データからルールを見つける	人がルールを決める
メリット	データの特徴に従い、アルゴリズムによって何らかの数学的根拠をもとにルールを見つける	<ul style="list-style-type: none">・ 自明なルールを決めることが非常に簡単・ ルールを複数組み合わせることで複雑なルールを作ることも可能
デメリット	<ul style="list-style-type: none">・ 一定以上のデータがないと識別や予測に効果的なルールを見つけられない・ 課題やデータに対して適切なモデルやアルゴリズムを選択する必要がある	<ul style="list-style-type: none">・ ルールのメンテナンスが大変・ 人が認識している以上の複雑で精緻なルールは作れない

▶ 構造化データ・非構造化データの例 図表15-1

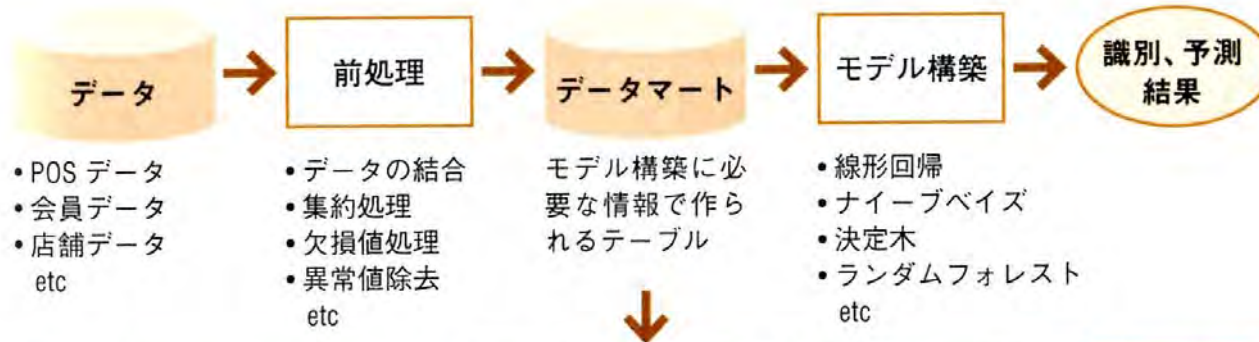
構造化データ

- 各種業務システム内のデータ(受注、発注、在庫、人事、POS など)
- 政府や調査会社の統計データ

非構造化データ

- 画像データ (商品画像、SNS の投稿画像など)
- 動画データ (監視カメラ映像、テレビ番組など)
- テキストデータ (議事録、SNS 投稿テキストなど)
- 音声データ (コールセンター会話録、会議録音データなど)

行と列を持つ表形式のデータ



顧客ID	年齢	性別 (男性:1,女性:0)	来店回数	最終来店後 経過日数	...	商品A購入
1001	40	1	11	35	...	1
1002	29	0	5	12	...	0
1003	48	1	...	50	...	1
...
1009	3	0
1010	18	0	12	10	...	1
...

説明変数
(パターン、ルールを見つけるもとのデータ)

目的変数
(識別、予測の
対象データ)

非構造化データ: 画像データに対する機械学習

画像はピクセルという小さなデータの集合体で構成されている
カラー画像ならRGBのそれぞれの数値としてピクセルに記録されており、たとえば、640X320ピクセルのカラー画像は、横に640ピクセル、縦に320ピクセルの画像となる。ピクセルごとに3色分、全部で614,400個の情報で画像が構成されている。ディープラーニングでは、画像のピクセルごと情報を全て入力として学習、推論を行う。

「画像分類」: 画像に写っている物体を特定のカテゴリに分類するタスク

「物体検出」: 特定の物体が画像内のどこに写っているかを検出するタスク

「シーン理解」: 画像がどのような状況を意味しているかを識別するタスク

「シーン理解」タスクが、最も高度で難しい。・・・自動運転などへの適用

非構造化データ：自然言語データに対する機械学習

他のデータ形式と同様に、文章を数値的なデータとして表現する必要がある。代表的な表現方法は「Bag of Words」と呼ばれるもので、文章ごとに単語の出現数をカウントし、文章と対応する単語の要素を列として表形式で表現する。

英語では単純であるが、日本語は複雑で難しい。

元の文章

これはペンです

あれは車です

それは鉛筆です

形態素解析による分割

これ／は／ペン／です

「これはペンです」という文章を Bag of Wordsによりベクトル化した例。文章に出現する「これ」「ペン」「です」に数字の「1」が入り、ほかの文章に含まれる単語はカウントされず「0」になる。このようにして数値の羅列に変換することをベクトル化と呼ぶ

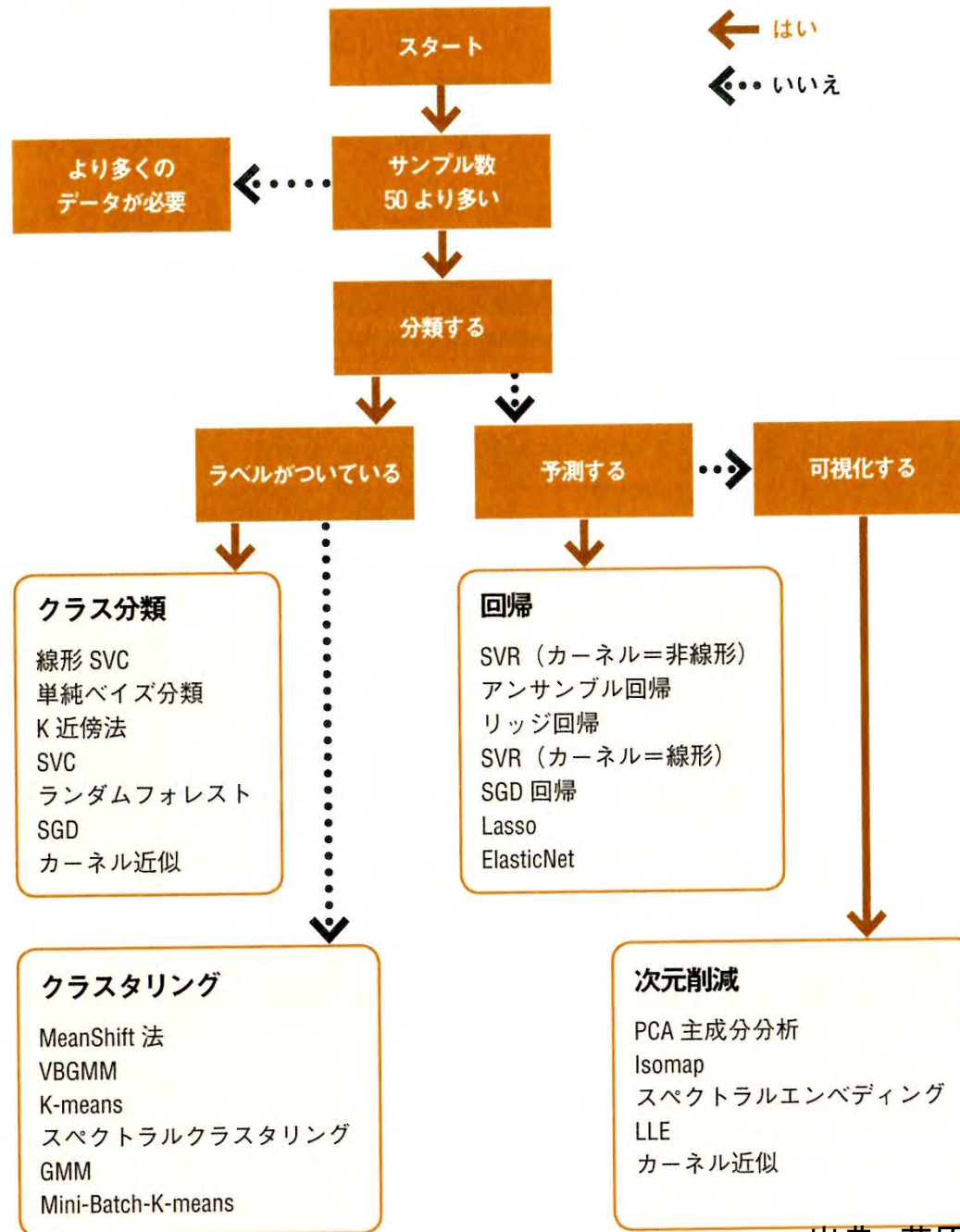
ベクトル化

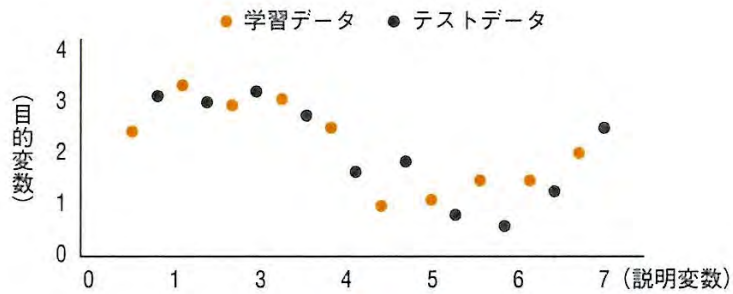
これ	あれ	それ	...	鉛筆	ペン	車	...	です	ない
1	0	0		0	1	0		1	0

形態素解析の例

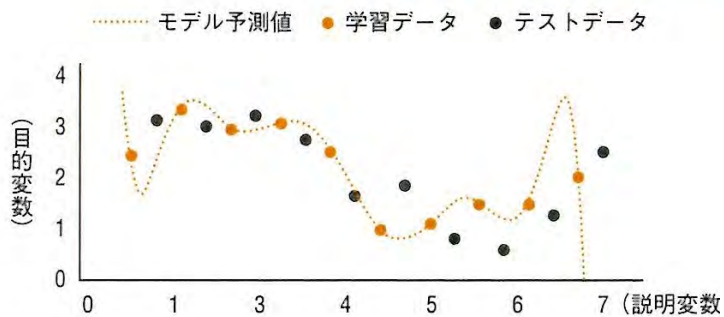
出典 荻原祐介:機械学習プロジェクトの教本

▶ アルゴリズムの選び方 図表50-3





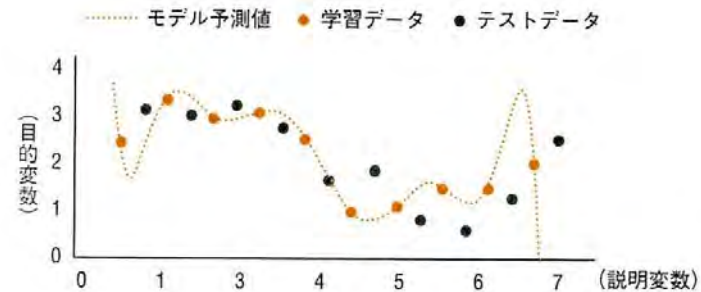
↓ 複雑なモデルを適用した場合



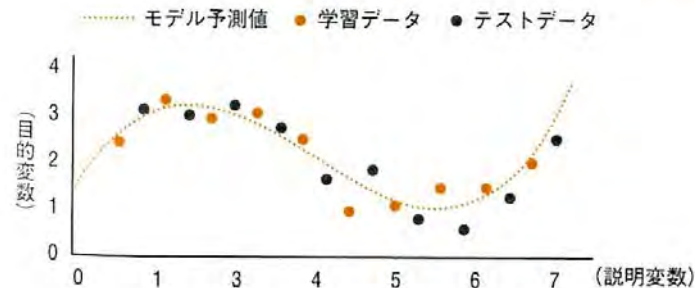
過学習について:

学習データに対してのみ適合した学習を行ってしまうこと

- ・学習データに対しては分類や予測の精度が高いが、別のデータに適用してみると、学習データに示した程度の精度を示さない状況



↓ ハイパーパラメーターを調整することでモデルを単純にする



モデルを単純化したことでテストデータの精度が大きく向上している
= 過学習が改善されている

ニューラルネットワークとディープラーニング

ディープラーニングとは、ニューラルネットワークの層を多数重ねたもの

手法	説明
ディープニューラルネットワーク (Deep Neural Network : DNN)	ニューラルネットワークの層を多層にしたモデル。一般的には中間層が2層以上の深い構造を持つ
オートエンコーダ (AutoEncoder)	ニューラルネットワークの出力が入力と一致するように構成したモデルであり、主に次元削減のために利用される
畳み込みニューラルネットワーク (Convolutional Neural Network : CNN)	主に画像解析に用いられるモデルで、局所的な領域の情報を畳み込む処理を行う畳み込み層と特徴を集約するプーリング層を多層に組み合わせて構成される
リカレントニューラルネットワーク (Recurrent Neural Network : RNN)	主にテキスト解析や時系列解析に用いられるモデルで、隠れ層の値を再び次の層の入力として用いるという再帰的な構造を持つ

さまざまなディープラーニング関連のモデル

プログラミング言語: 代表的なもの Python

機械学習に必要な数値計算を高速で行う「NumPy」、機械学習のライブラリーであるのライブラリーである「scikit-learn」、ブラウザ上でインタラクティブにプログラミングを進めることができる「Jupyter Notebook」、Googleのオープンソース「Tensor Flow」が利用できる

▶ Pythonの機械学習用ライブラリとツール 図表22-2

ライブラリ/ツール名	説明
NumPy	機械学習に必要なベクトルや行列の計算を高速に処理するライブラリ
Pandas	データ分析に必要なさまざまな機能が提供されているライブラリ
scikit-learn	多数の機械学習アルゴリズムを提供しているライブラリ
Jupyter Notebook	ブラウザ上でインタラクティブにデータ分析や機械学習に取り組めるツール
TensorFlow	Googleがオープンソースとして公開した機械学習/ディープラーニングのためのライブラリ

主なオープンソースの機械学習のライブラリ／言語

ライブラリ／言語	開発組織名	説明
scikit-learn	scikit-learn project	2007年にスタートしたオープンソース機械学習ライブラリ。操作言語はPython
TensorFlow	Google	2016年11月に公開された機械学習ソフトウェアライブラリ。Google社内で機械学習やディープニューラルネットワークの研究にあたるメンバーによって開発され、Google社内で使われていたものがオープンソース化された。操作言語はPython。C++、Java、Goは試験利用段階
Chainer	Preferred Networks	ニューラルネットワークのデザイン、学習、評価といったディープラーニングのモデル構築をサポートする、オープンソースのフレームワーク。操作言語はPython
Caffe	Berkely Vision and Learning Center	Googleに在籍するエンジニアがカリフォルニア大学バークレー校の博士課程在籍時に開発を開始し、オープンソースで提供されているディープラーニングのフレームワーク。操作言語はPython
CNTK	Microsoft	ディープラーニングのためのツールキット (Cognitive toolkit : CNTK)。操作言語はPython、C#
R言語	R Foundation	統計解析のためのプログラミング言語およびその実行開発環境で、さまざまな機械学習アルゴリズムのパッケージが利用できる

▶ ストレージとGPU 図表25-1



出典：CiscoのWebページより
ストレージの例。写真はCisco UCS S ストレージサーバ

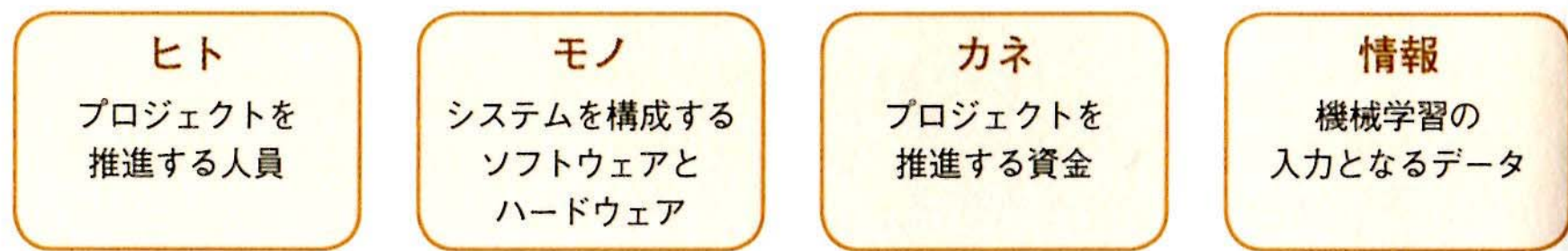


出典：NvidiaのWebページより
GPUの例。写真はNvidia Tesla GP

機械学習に必要なストレージと計算資源

出典 荏原祐介：機械学習プロジェクトの教本

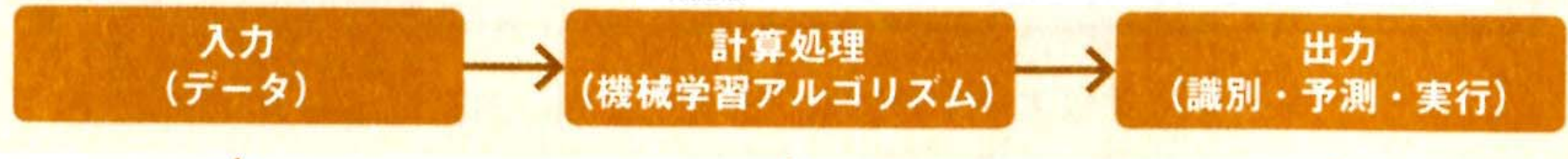
▶ 機械学習プロジェクトに必要なリソース 図表20-1



	構想フェーズ	PoCフェーズ	実装フェーズ	運用フェーズ
(期間の下限)	1か月～	1か月～	3か月～	—
(参画人数規模の下限目安)	1.5人/月～	1.5人/月	3人/月	1人/月

ソフトウェア：機械学習の処理の実装 (参画人数規模の下限目安)

	構想フェーズ	PoCフェーズ	実装フェーズ	運用フェーズ
約330万円～	約330万円～	約1,980万円～	約220万円/月	



ハードウェア：機械学習の処理を支えるインフラ



構造化データの質

そもそものデータの
間違い、カラムのズレ、
数値の丸め誤差、さまざま

▶ データ間違いの例 図表49-2

<正しいデータ>

	Sun	Mon	Tue	Wed	Thu	Fri	Sat
A	0.089	0.092	0.102	0.362	0.118	0.098	0.124
B	0.533	0.612	0.565	0.432	0.823	0.998	0.984
C	0.902	1.102	1.332	0.805	1.201	1.109	0.856

<カラムのずれ、欠損>

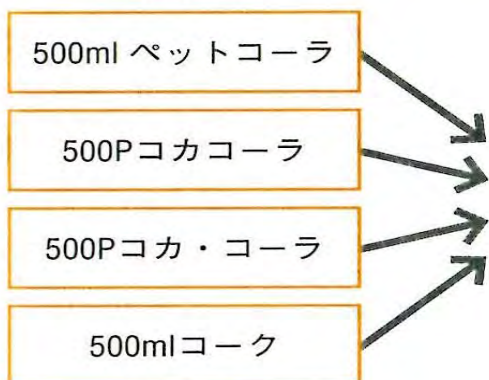
		Mon	Tue	Wed	Thu	Fri	Sat
A	0.089	0.092		0.362	0.118	0.098	0.124
B	0.533	0.612	0.565	0.432	0.823		0.984
C		0.902	1.102	0.565	0.805	1.20	1.201

<数値の丸め>

	Sun	Mon	Tue	Wed	Thu	Fri	Sat
A	0	0	0	0	0	0	0
B	1	1	1	0	1	1	1
C	1	1	1	1	1	1	1

▶ 名寄せのイメージ 図表50-2

表記揺れ



同じデータでありながら表現が異なるもの

名寄せ

商品	容器	容量 (ml)
コカ・コーラ	ペットボトル	500

前処理の必要性

機械学習においては、モデル構築が最も大変と思われがちですが、多くの場合、前処理の工数に最も時間がかかる。モデル構築は、学習アルゴリズムを適用するだけで済むが、前処理は問題によって適切な加工方法が異なるためである。

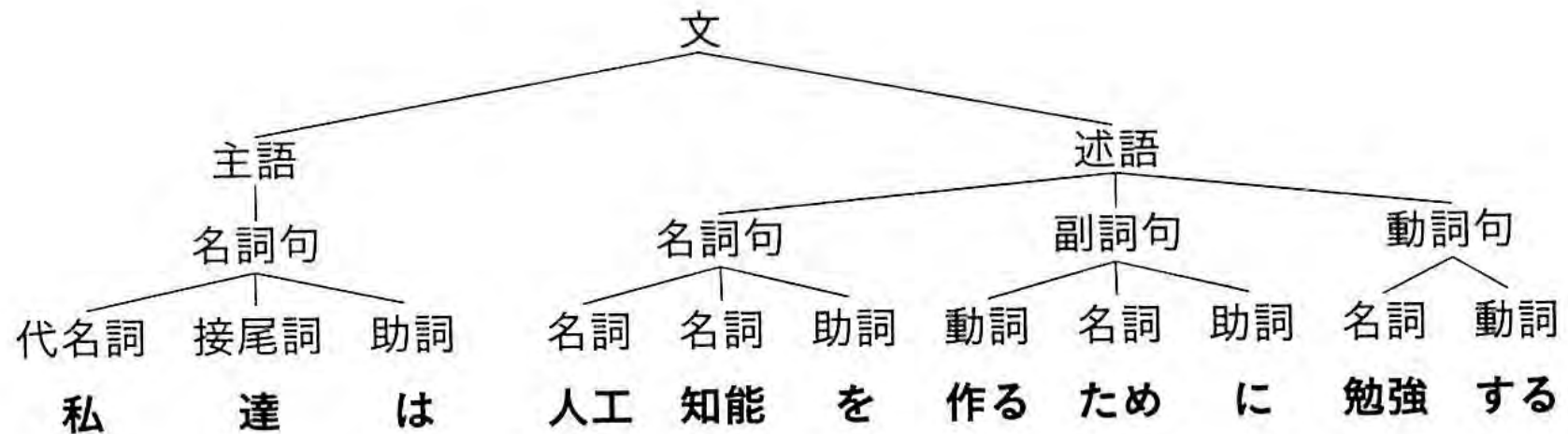
画像処理の場合、識別に用いる領域を切り抜く処理や大きさを揃える処理が必要となる。

自然言語の場合は、適切な形態素解析を行うための「辞書」と呼ばれる単語リストが必要となる。

実践編 人工知能が実際に使われている応用例

自然言語処理 たとえば 音声認識、言語の翻訳など

- ✓ 自然言語処理に、数学的にどういった挑戦があるかを理解する。
- ✓ 線形代数（ベクトル，行列）と自然言語の関係を理解する。



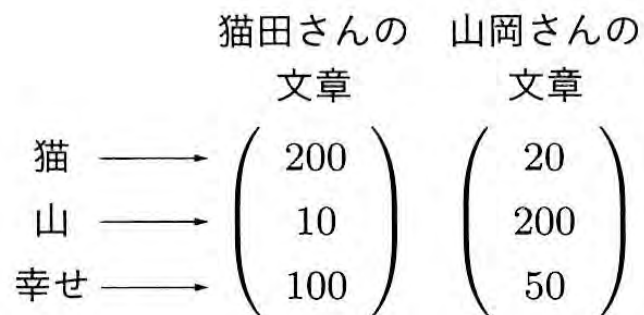


図 6.6.1 ベクトル化した文章

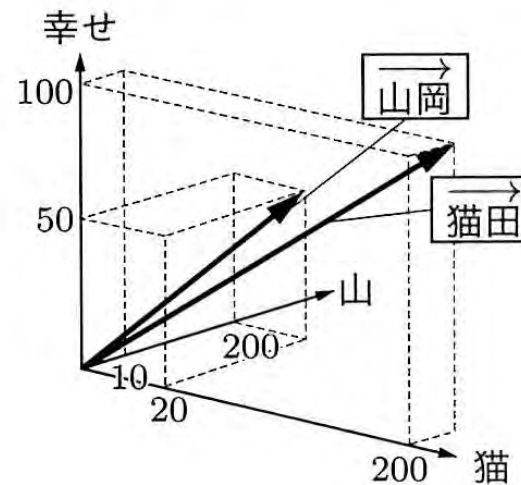
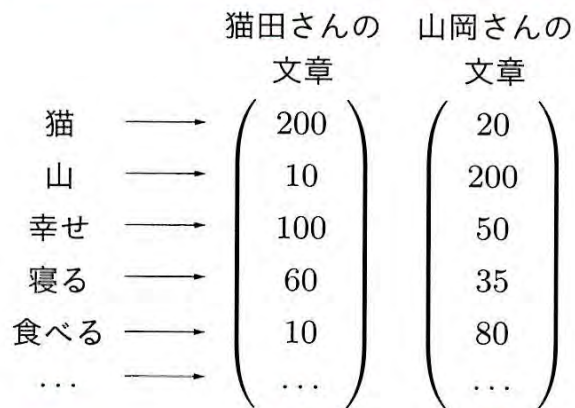


図 6.6.2 ベクトルの図示

「猫」「山」「幸せ」「寝る」「食べる」…
辞書に含まれる合計 40 万語



↓ ↓ ↓
最大 40 万次元の Bag-of-Words ベクトル

図 6.6.3 ベクトル表現の方法 出典 荻原祐介:機械学習プロジェクトの教本