

記 録

文書番号	S C J 第 23 期-290915-23530300-066
委員会等名	日本学術会議 情報学委員会 e-サイエンス・データ中心科学分科会
標題	ビッグデータ利活用展望調査
作成日	平成 29 年（2017 年）9 月 15 日

※ 本資料は、日本学術会議会則第二条に定める意思の表出ではない。掲載されたデータ等には、確認を要するものが含まれる可能性がある。

この記録は、日本学術会議 情報学委員会 e-サイエンス・データ中心科学分科会の審議の記録を公表するものである。

日本学術会議情報学委員会 e-サイエンス・データ中心科学分科会

委員長	北川源四郎	(第三部会員)	明治大学先端数理科学インスティテュート所員
副委員長	樋口 知之	(連携会員)	情報・システム研究機構理事、統計数理研究所長
幹 事	高安美佐子	(連携会員)	東京工業大学大学院総合理工学研究科准教授
幹 事	鷲尾 隆	(連携会員)	大阪大学産業科学研究所教授
	大倉 典子	(第三部会員)	芝浦工業大学工学部情報工学科教授
	相澤 彰子	(連携会員)	国立情報学研究所コンテンツ科学研究系教授
	相田美砂子	(連携会員)	広島大学大学院理学研究科教授
	新井 紀子	(連携会員)	国立情報学研究所情報社会相関研究系教授、社会共有知研究センター長
	有田 正規	(連携会員)	国立遺伝学研究所教授
	今井 桂子	(連携会員)	中央大学理工学部教授
	岩野 和生	(連携会員)	株式会社三菱ケミカルホールディングス先端技術・事業開発室執行役員
	上田 修功	(連携会員)	NTT コミュニケーション科学基礎研究所 NTT フェロー、機械学習・データ科学センター代表
	加藤 直樹	(連携会員)	関西学院大学理工学部情報科学科教授
	喜多 泰代	(連携会員)	国立研究開発法人産業技術総合研究所知能システム研究部門上級主任研究員
	小林 広明	(連携会員)	東北大学大学院情報科学研究科教授、サイバーサイエンスセンターセンター長特別補佐
	椿 広計	(連携会員)	独立行政法人統計センター理事長
	徳山 豪	(連携会員)	東北大学大学院情報科学研究科教授
	中島 秀之	(連携会員)	東京大学大学院情報理工学系研究科特任教授
	西尾章治郎	(連携会員)	大阪大学総長
	松井 知子	(連携会員)	統計数理研究所モデリング研究系教授
	松田三知子	(連携会員)	神奈川工科大学情報学部教授
	松本 裕治	(連携会員)	奈良先端科学技術大学院大学教授
	安信千津子	(連携会員)	株式会社日立製作所情報・通信システム社経営戦略室企画本部技術戦略ユニット
	山西 健司	(連携会員)	東京大学大学院情報理工学系研究科数理情報学

		専攻教授
渡辺美智子	(連携会員)	慶應義塾大学大学院健康マネジメント研究科教授

本記録の作成にあたり、以下の職員が事務を担当した。

事務	石井 康彦	参事官（審議第二担当）（平成 29 年 7 月まで）
	糸川 泰一	参事官（審議第二担当）（平成 29 年 7 月から）
	松宮 志麻	参事官（審議第二担当）付参事官補佐（平成 29 年 7 月まで）
	高橋 和也	参事官（審議第二担当）付参事官補佐（平成 29 年 7 月から）
	水野 雅弘	参事官（審議第二担当）付審議専門職付（平成 28 年 3 月まで）
	駒木 大助	参事官（審議第二担当）付審議専門職付（平成 28 年 4 月から）

1. はじめに

(1) 展望調査の背景と趣旨

情報通信技術、特にセンサーの飛躍的発展とユビキタス化によって多くの学術研究分野や社会の隅々においてビッグデータが出現し、歴史の転換点をなすエポックメイキングな活用例をしばしば見聞きするようになっている。ただし、ビッグデータには膨大な知識や潜在的価値が埋蔵されていることが期待されるとはいえ、その有効活用には高度な解析技術や計算基盤が要求されるため、ビッグデータ利活用に向けて、研究課題、研究基盤、社会制度、法制度などの変革を伴う激しい国際競争が始まっている。我が国においても、将来を見通した基盤技術の開発とデータ整備および具体的な現場でのデータ活用を積極的に推進しなければ、新たな局面を迎えた学術研究や産業界において国際的に大きく立ち遅れる恐れがある。

当分科会が2014年に発出した提言「ビッグデータ時代に対応する人材の育成」[1]では、ビッグデータ活用に必要な要素技術として、ビッグデータ処理技術、データ可視化、データ解析法の3つを挙げ、これらの発展の結果として描かれるビッグデータが拓く新しい社会の姿としては、

- ビッグデータの活用による個人化サービスの実現、地球規模での網羅的データを活用したデータ駆動型産業の創出。
- 多数のセンサーからの情報や先進的シミュレーション技術を活用した1次産業・2次産業の効率化。
- ゲノム情報や医療情報の統合活用による個別化医療や予防医学の実現。
- 大量配置のセンサー情報やシミュレーション技術の統合による、社会インフラのスマート化。
- 公共投資、観光政策、環境対策などにおける、データに基づく科学的意思決定・政策決定の実現。
- ビッグデータ活用により、従来の手法では見過ごされてきた稀少事象やリスクの発見による安全社会の実現やイノベーションの創出。
- モバイル情報や個人・集団の位置情報等を活用した災害時の効果的対応。
- ビッグデータ活用による人文学の新しい方法論の開拓。

などを挙げた。

しかし、その後もスパースモデリング*、深層学習などの応用が驚くべく速さで進んでおり、コンピュータ碁のプロ棋士への完勝、自動翻訳や自動運転技術の飛躍的発展、ドローンの活用、防犯カメラや車載カメラの活用などはマスコミにもしばしば取り上げられたように、一般社会に情報科学技術発展のインパクトを印象付けた例は枚挙にいとまがない。

このような状況にあたって、当分科会では、前回の提言作成時の議論を踏まえつつ

も、その後3年が経過した現時点において技術的な観点からビッグデータの利活用について、改めてその定性的あるいは定量的発展のレベルやその変化傾向について検討を行うとともに、その発展動向がもたらす社会的波及効果の抽出を試みた。

分科会では更に、5年後、10年後における発展動向とその考えられるインパクトの予測を試みた。もとより、情報科学技術は日進月歩であり、それまでの最先端技術が一夜にして陳腐化してしまった例にも見られるように、長期的展望は極めて困難である。それにも関わらず、あえて10年先までの技術予測を試みたのは、基礎研究においては10年以上先を見通した長期的課題設定が重要であることと、技術開発においては経験則にすぎない「半導体の集積率は18か月で2倍になる」というムーアの法則*に従うかのごとく、半導体性能の技術発展が進んできたことに見られるように、社会の予想や期待が技術発展をリードするという側面があるからである。

（２）展望調査で取り上げた技術課題

展望調査では下記の12の技術課題に対して、それぞれ発展動向の観点を複数定め、それぞれの軸に沿っての発展予測を試みた。

1. 計算基盤
2. ビッグデータ、データベース
3. シミュレーション・データ同化
4. 機械学習、深層学習
5. パターン認識（画像認識、物体認識、音声認識）
6. 自然言語処理
7. ロボット
8. マニファクチャリング
9. ネットワーキング
10. フィンテック・電子契約
11. 行政マイクロ情報等の利活用
12. 人工知能リスク管理

これらを大雑把に分類すると、1、2は計算基盤に関する課題、3－6はビッグデータ利活用のための方法的基盤、7－12は具体的な利活用に関連した課題である。

なお、上記12課題以外にもスパースモデリングや最適化などの重要な技術課題もあるが、分科会の活動の記録という性格上、十分な議論を行なえなかった項目は取り上げていない。したがって、本展望調査はビッグデータ利活用のための重要な技術課題を取り上げてはいるが、必ずしも網羅的に取り上げたものではないことは注意しておきたい。

発展動向の観点としては、技術の高度化、機能の高度化、対象や応用の多様化・大規模化などが比較的共通に考えられるが、ここでは技術課題ごとに3つ程度の具体的な観点を定め、その観点から現在の動向、5年後の動向、10年後の動向の検討を行った。次

節以降では、上記 12 の技術課題についてそれぞれの検討結果を列挙し、分科会での議論を記録に留める。

なお、わかりにくい専門用語については本記録の最後に用語解説を述べ、各々初出の部分に*印を付した。

2. 各技術課題の発展動向

(1) 計算基盤の発展動向

計算基盤とは、超並列処理など最新の計算技術を用いて大規模シミュレーションや大規模データ処理を行うための基盤となる高性能計算機システムのことである。ここでは、その発展動向を3つの観点：計算基盤の高性能化の発展を示す「性能向上」軸、AI・ビッグデータを利活用するための計算基盤(アーキテクチャ)の高度化の発展「AI・ビッグデータ向けアーキテクチャの高度化」軸、そしてIoT(Internet of Things、もののインターネット化)などセンサーと連携した計算基盤全体の構築技術の発展に関する「計算基盤構築技術」軸に沿って整理する。

現在の動向

「性能向上」に関しては、計算科学分野でのHPC(高性能計算)技術の発展を受けて、超並列処理やベクトル処理を大規模データ解析に活用したAI・ビッグデータ計算基盤が一般化しつつある。特に、単純な演算器を大量に集積し高性能化を実現しているグラフィックスプロセッサを汎用処理に用いるGPGPU(General-Purpose computing on Graphics Processing Units)のHPC分野での成功は、その応用範囲を従来のシミュレーション科学分野における計算中心のアプリケーション開発だけでなく、データサイエンス分野のデータ中心のアプリケーション開発へと大きく展開されつつある。

「AI・ビッグデータ向けアーキテクチャの高度化」については、従来のノイマン型計算モデルをベースに構築された超並列計算モデルに、次々に生み出される様々なAI・ビッグデータ利活用のためのアルゴリズムを実装しているのが現状である。その結果、計算中心のコンピュータアーキテクチャをうまく活用できるアルゴリズムしか成功しておらず、大量のデータを扱うビッグデータ解析を従来の計算中心のコンピュータアーキテクチャに適用した場合、メモリやファイルI/Oがボトルネックになって、必ずしも期待される性能が引き出せないような状況が生じている。このことがAI・ビッグデータ技術のさらなる進展を制約してしまう懸念が生じつつある。「計算基盤構築技術」に関しては、IoTセンサー群をスタンドアロンコンピュータやサーバに接続し、集中処理しているのが現状である。従って、今後ますます増え続けるデータを効率よく処理し、そこから更なる高度な情報を引き出すためには、サーバ側の負荷を減らすことが必要であり、小規模サーバの分散配置(いわゆるエッジコンピューティング*)やIoT機器自身でのデータ処理を行う等、IoT側(より利用者に近い側)でデータの選別を行うなどの1次処理が求められてくると思われる。

5年後の動向

「性能向上」に関しては、今後データサイエンス分野の発展に伴い、現在の計算性

能重視(flop/s 指向*)のアーキテクチャ(Compute-Centric Architecture)設計から、大規模データ解析に必要なファイルI/Oからメモリまで一定時間内に処理できるデータ量を重視するデータスループット指向*(Bytes/s 指向*)アーキテクチャや大規模メモリ重視のメモリ中心アーキテクチャ(Memory-Centric Architecture)設計へのパラダイムシフトが起こると思われる。もちろん、従来のアーキテクチャが衰退するのではなく、様々な応用領域に適したアーキテクチャが開発され、その分野のアプリケーションに適した計算基盤の性能最大化が得られるようになる。このことは、「AI・ビッグデータ向けアーキテクチャの高度化」とも深く関係し、AI・ビッグデータ利活用においても様々なアルゴリズムが考案されていることから、それらの特徴を考慮し、それぞれに適したプラットフォームの実現と AI・ビッグデータ共通計算基盤としてのそれらの抽象化技術の発展が強く望まれる。「計算基盤構築技術」に関しては、IoT のインテリジェント化 (AI of IoT) の発展により、エッジでのデータのフィルタリングとメタデータ化 (情報トリアージ技術*)、サーバ側での高次データ解析技術とその AI データベース構築技術が発展すると思われる。

10 年後の動向

「性能向上」に関しては、3次元実装・シリコンフォトリソグラフィ等必ずしもムーアの法則に従わずいろいろな機能をデバイスに盛り込んだ More-Than-Moore 技術*を活用して、レジスタ・キャッシュ・メインメモリ・ファイルシステムを統合的に扱うユニバーサルメモリアーキテクチャとその各階層のインテリジェント化のためのインメモリコンピューティングアーキテクチャ*が確立し、データサイエンス向け高性能計算基盤が構築されることが予想される。また、「AI・ビッグデータ向けアーキテクチャの高度化」に関しては、AI・ビッグデータのための抽象計算モデルを実現する「マイクロアーキテクチャ」設計のために、マルチコア・メニーコア、FPGA (field-programmable gate array)、アルゴリズム適応型精度保証可変長計算、オンチップ動的な可変ネットワーク、ユニバーサルメモリ、インメモリコンピューティングなどを More-Than-Moore 技術で従来の半導体の集積率が 18 か月で2倍になるという Moore の法則を前提とした Moore 型*マイクロアーキテクチャを高度化し、これらとニューラルネットワークに代表される Brain-Inspired Computer など Moore の法則が適用できなくなることを前提とした Post-Moore 型*アーキテクチャが融合した新しいアーキテクチャが実用化されることが期待される。これにより、「演繹処理」と「帰納処理」の同一アーキテクチャ上での統合処理が実現され、新しい AI・ビッグデータ処理基盤として発展していくことが期待される。「計算基盤構築技術」に関しては、仮想マシン上に構築される AI データベースのオープンデータ化とインタークラウド技術による AI 連携 (分散共有処理) の発展により、サイバーフィジカルシステム*(実世界 (フィジカル空間) にある機器やインフラなどに関する多様なデータをセンサーネットワーク等で収集し、サイバー空間で大規模データ処理技術等を駆使して分析／知識

化を行い、そこで生み出された情報によって、それらの障害を未然に防いだり最適制御をおこなうことを可能にするもの)が現実のものとなることが予想される。

(2) ビッグデータ、データベースの発展動向

ビッグデータとは、既存の一般的な技術では管理することが困難なほど大量あるいは複雑なデータのことである。より広義には、多様性、データ量、データの生成頻度、更新頻度の点で管理が困難なデータ、および、それらを蓄積・処理・分析するための技術を含む概念である。データの巨大さゆえ、「ビッグデータアルゴリズム」の開発と、データ圧縮技術に関連する「リアルタイム処理技術の高度化」が、重要課題である。データベースとは、検索や蓄積が容易にできるよう整理されたデータの集まりのことである。データベースへのデータの格納方法を論理的・物理的に規定するデータモデルには、階層型データモデル、リレーショナルデータモデル、オブジェクトデータモデルなど、いくつかの代表的モデルが知られている。ビッグデータと関連したデータベース技術の高度化が重要課題である。本稿では、「ビッグデータアルゴリズム」、「リアルタイム処理技術の高度化」、「ビッグデータに対するデータベース技術の高度化」の3つを発展動向の観点とする。

現在の動向

「ビッグデータアルゴリズム」については、スーパーコンピュータを用いて数10億点のグラフに対する探索や最短路計算が高速にできるようになっている。また、データのランダムサンプリング技術がグラフの特性量の推定や最適化問題に適用できることが明らかになり、定数時間アルゴリズムが開発された。これにより、データサイズによらず、定数時間で、大規模な問題も近似的に解くことができる。しかし、まだ理論的段階なので、実用化のための技術開発が必要である。「リアルタイム処理技術の高度化」については、センサーデータなどリアルタイムで発生するビッグデータを圧縮しながら蓄積、処理するストリームアルゴリズムが実用化されつつある。データロスの無い高速伝送が必要な状況が発生する医療画像データの伝送、車載ネットワークなどへの応用が期待される。「ビッグデータに対するデータベース技術の高度化」については、ビッグデータを大規模分散環境(たとえば、分散コンピューティングを支援するプログラミングモデルである MapReduce)の下で、数時間～一昼夜程度の計算時間でデータ解析可能になった。オンメモリ分散処理(たとえば、Spark*)による準リアルタイム処理も視野に入っている。

5年後の動向

「ビッグデータアルゴリズム」については、定数時間アルゴリズムの実際的評価が幅広く行われ、より広範な問題群を定数時間で解くための方法論が整備され、多くの

問題に対して、データサイズに依存せず、近似的に定数時間で、解を生成する乱択（確率的）アルゴリズムが実用化される。「リアルタイム処理技術の高度化」については、高速伝送を実現するストリームアルゴリズムが、さまざまな機器に組み込まれ、交通、経営、医療の分野での実用化が実現している。「ビッグデータに対するデータベース技術の高度化」については、センサーデータやIoTからのセンサーストリームと蓄積されたビッグデータが融合し、機械学習・ビッグデータ解析が可能になる。データの品質（精度、信頼度、フォーマット、データリネージ*、プライバシー）をコントロールするための技術が実用化され始めることが見込まれる。

10年後の動向

「ビッグデータアルゴリズム」については、定数時間アルゴリズムは、ビッグデータを対象とする次世代の新しい計算モデルとして世の中に定着し、例えば、超大規模なグラフに対する最短路計算などのさまざまなアルゴリズムを近似的に解く方法が実用化され、諸分野へ応用されていることが見込まれる。「リアルタイム処理技術の高度化」については、遠く離れた観測装置から得られる観測データを世界各地に高速伝送し、共有しながら解析ができるなど、科学分野での活用が行われることが見込まれる。「ビッグデータに対するデータベース技術の高度化」については、蓄積データやセンサーデータを覆う統一的な情報空間の上で、データの品質（精度、信頼度、フォーマット、リネージ、プライバシー）がコントロールされた異種データが、解析者の必要に応じて高い自由度で統合され、機械学習・ビッグデータ解析にシームレスに結合されることが、見込まれる。

（3）シミュレーション・データ同化の発展動向

シミュレーションとは狭義には、研究対象の現象を記述する支配方程式を空間・時間に関して離散化し、その時間発展あるいは空間変形（伝搬）、またその両方を大規模数値（積分）計算で解くものを指す。離散化された計算式をシミュレーションモデルと呼ぶ。一方、対象に関する支配方程式が存在しなくとも、解析者の直感や当該分野の経験にもとづく数理モデルを採用し、狭義同様に、大規模数値計算で時間発展・空間変形を解くものが広義のシミュレーションである。ここでは広義のシミュレーションを考える。シミュレーションモデルは通常未確定のパラメータを含み、さらに実問題のシミュレーションを行う際には、初期条件や境界条件も問題毎に定めなくてはならない。このようにシミュレーションモデルには数多くの多様な不確実性が存在するため、シミュレーションを活用して予測や判断を行うには、不確実性をなるべく減ずることが必須である。その一策が、シミュレーションの結果を実際のデータとなるべく合うように、シミュレーション内のパラメータ等を調整する作業、つまりデータ同化*になる。以下では、研究発展の動向を「計算規模」、「ソフトウェアのオープン化」、

「可視化や結果解釈機能の高度化」の3つの観点から考察する。

現在の動向

近年のビッグデータの登場により、データ同化のコンセプトと適用範囲はより広範な分野に広まり、シミュレーションを積極的に活用していた分野でその有用性が深く認識されつつある。具体的には、さまざまな流体现象を取り扱う流体シミュレーション、量子力学にもとづく物質科学・材料開発、人工物の熱特性を解析する熱シミュレーション等である。「計算規模」に関しては、シミュレーション単体では現在、数千万から10億程度の変数を含む問題を、またデータ同化では1000万程度の変数に対する同化計算が可能である。「ソフトウェアの高度化」に関しては、空間自動メッシュ化がほぼ実現、またダウンスケーリング（時間・空間解像度の超高精度化）については静的方法が実現されている。「可視化や結果解釈機能の高度化」に関しては、探索的ではあるが、必要な分析観点の選択が半自動的になりつつある。

5年後の動向

シミュレーションと観測・計測技術の両者は研究開発現場でも別に行われがちであったこともあり、それらの統合効果には大きな期待が寄せられている。特に、観測や計測に長い時間や多大な予算がかかる場合、効率的な観測・実験の実施による開発予算の削減である。これらは、スマートシミュレーション、あるいはスマートデザインといった言葉で語られるような統合技術で、今後5年間に研究が大いに進むであろう。大量のシミュレーションの結果を観測・計測データとして取り扱い、ターゲットとするシミュレーション変数を機械学習により統計モデル（例えば回帰式）で表現する、いわゆるエミュレーション技術も、スマートシミュレーションのサブ課題として重要になる。「計算規模」の面では、シミュレーション単体で現状の10倍（解像度で言えば1/10）程度、データ同化計算の規模は現状の1000万程度を大きく上回ることは難しいと予想する。「ソフトウェアの高度化」に関しては、ダウンスケーリングの自己無撞着性が向上し、シミュレーション内のマクロパラメータの自動同定も実現される。「可視化や結果解釈機能の高度化」については、人の対象理解を促進する上で効果的な変数（特徴量・アспект）の選択がスパースモデリングおよび深層学習により自動化されるであろう。位相データ解析の進歩と連動した効果的可視化も実現されると予想する。

10年後の動向

大規模な人工物、例えば社会インフラからの膨大なセンサー情報を適切に利用して社会インフラを常時モニターし、必要なメンテナンスを素早く施すことにデータ同化技術は直接的に役に立つ。また、生体に非侵襲の計測装置*から得られる情報の領域や精度には限界があるが、データ同化によりそれらの限界を突破することが可能である。

その結果、術前に個人の臓器形や機能にあわせたシミュレータを用意し、常に複数のシナリオを先読みしながら手術を実施することも可能になろう。「計算規模」はシミュレーション単体で現況の100倍程度、データ同化は1億程度の変数に対して同化計算が可能になると予想する。「ソフトウェアの高度化」については、個人/グループレベルで開発するシミュレーションモデルと既存のモデル群とを連動させるエコシステムが実現する。「可視化や結果解釈機能の高度化」については、シミュレーションコードと結果の因果解析を総合的に活用することで、因果構造を半自動的に抽出する機能を実現されているであろう。

(4) 機械学習・深層学習の発展動向

機械学習とは、人間の学習能力と同様の機能を、個別の学習問題について人が明示的にプログラミングすることなく汎用的にコンピュータで実現しようとする技術・手法のことであり、具体的には、対象とするデータ間の特徴的關係やパターンを抽出し、それらを用いて対象と同じあるいは類似の機構から生成された新たなデータについて識別・分類やクラスタリング、異常検知を含む推定やそれらの予測を行う。また、深層学習とは、多層のニューラルネットワークを用いた機械学習技術の総称であるが、特に層数が3を超えるものを指すことが多い。現状から10年後までの発展動向においては、以下の3点が社会実装の面で特に重要な変化を引き起こすと考えられる。「対象データの大規模化」：事例数や属性次元数、関係の複雑さ・多様性など、様々な面での対象データの規模。「学習の自律性」：アクティブな事例・属性選択、強化学習のような自律的な事例・属性を収集、自律的な学習目的や基準の選択・設定。「対象問題の多様化」：識別・分類やクラスタリング、外れ値検出などの典型的機械学習問題に加えて、情報検索やシミュレーション、計測、意思決定など他の情報処理問題を扱う理論や技術と、その融合による多様化。

現在の動向

「対象データの大規模化」については、現状、深層学習などで時空間的構造を持つ大量データからの高精度学習が実現されているが、膨大な学習データが必要である。「学習の自律性」については、強化学習で複雑な状態や文脈に基づいて、効果的な自律的情報収集・試行を行う学習技術が実現されており、さらに深層学習を導入して大規模・複雑な学習を自律的に行えるようになりつつある。「対象問題の多様化」については、予兆発見・検知や因果推論、クラウドソーシング技術、シミュレーションとの融合など、より幅広い対象問題に研究が広がりつつある。

5年後の動向

「対象データの大規模化」については、今後数年でシミュレーションを用いて仮学習し、実データに転移学習する方法が広く開発されると予想される。また、深層学習

を組み込むシステム化技術の発達により、大規模問題での学習実用化が加速し、さらに大規模問題に対するリアルタイム逐次学習技術が発展する。「学習の自律性」については、深層学習技術の発展によって現状よりさらに大規模・複雑な問題やデータに関する強化学習などが実現され、より広範な実問題への適用が進む。「対象問題の多様化」については、より多くの他の情報処理技術と機械学習技術の融合・分化が進み、多様な境界領域的研究やその技術が実現されていくと思われる。

10 年後の動向

「対象データの大規模化」については、10 年程度先には、時空間的構造を持たないデータに対しても高精度な学習が行える技術改良が進むか、あるいはデータを時空間的構造に埋め込んで高精度学習を行う技術が開発され、より多様なデータや問題への適用が進むと考えられる。「学習の自律性」については、10 年程度先には複雑・多様な問題構造を探索的に発見する技術の研究が進み、限定的だが問題定義の置き換えを行って、有限の情報処理能力や知識しか有しない人工知能には現実には起こりうる問題全てに対処することはできなというフレーム問題を緩和する学習技術が実現する可能性がある。「対象問題の多様化」については、異分野融合の大規模システムや複数分野のビッグデータを横断的に利用する機械学習技術が実現されていくと考えられる。

(5) パターン認識（画像認識、物体認識、音声認識）の発展動向

パターン認識とは、画像認識、物体認識、音声認識など、画像・音声などの雑多な情報を含むデータの中から、一定の規則や意味を持つ対象を選別して取り出す処理である。この処理では統計、機械学習技術が用いられ、その目的は識別や回帰に代表される。したがって、「機能の高度化」は識別や回帰の性能向上を指す。また現在、センシング技術の発展とともに多種多様なデータが利用できるようになっているが、「機能の多様化」とは各種データのいろいろな組み合わせによるパターン認識の実現、及び従来の識別や回帰を超えて、意思決定のための方策を自動獲得するなどの目的の多様化を含む。さらに、「機能の個性化」とは、各ユーザの特性に対応したパターン認識の実現を指す。

現在の動向

「機能の高度化」については、深層学習によりパターン認識の性能向上がはかられている。音声認識では数 1000 時間の学習データを用いた場合に誤り率 10%以下の精度[1]、一般物体認識では 1000 カテゴリについて誤り率 15%程度の精度[2]を得ることができるようになっている。また物体のクラスや識別に有効な画像・音声上の局所的な特報量（特徴量・アスペクト）が自動獲得されつつある。「機能の多様化」については、例えば、スパムや攻撃メールの検出においてパターン認識が利用されている。また道路気象や交通規制の情報を含め複数のデータを利用した交通量管理システム

が実現されつつある。さらに強化学習と組み合わせて、ロボット行動などの方策の自動獲得が実現されつつある。「機能の個性化」については、ユーザの話し方に対応した音声認識、ユーザ環境を考慮した物体認識などの個人化技術が実現されつつある。

5年後の動向

「機能の高度化」については、深層学習の機械学習技術が深化し、人手による調整が必要であったネットワーク構造やパラメータの初期値の設定が自動化して、パターン認識の性能およびインタフェースが向上すると考える。同時にパターン認識の応用領域の拡大が見込まれる。一般物体認識においては、物体のクラス間の構造が自動獲得されることが見込まれる。「機能の多様化」については、例えば、一般物体認識について、マルチモーダルデータ*の自然な統合によるパターン認識が実現されることが見込まれる。またパターン認識を利用した自動運転技術が普及することが見込まれる。さらに、高度なパターン認識と強化学習を用いた、各種エージェントの行動を決める報酬評価関数*の自動獲得が実現されることが見込まれる。「機能の個性化」については、パターン認識の各処理のユーザ対応が実現される。

10年後の動向

「機能の高度化」については、深層学習を超える統計的機械学習技術が開発され、パターン認識の性能およびインタフェースがさらに向上すると考える。いろいろな関係性（因果や階層構造等）が自動獲得されるようになると思う。また希少な事象のデータについても、高度なパターン認識と統計的機械学習を用いて解析できるようになると考えられる。さらに、自律的にデータ取得、学習を行い、パターン認識技術を自動で進化させる技術が研究開発されるようになると思う。「機能の多様化」については、例えば、パターン認識による一般物体認識システムは、進化するIoT技術等を用いて、自律的に認識対象を広げていくと考えられる。「機能の個性化」については、ユーザ対応したパターン認識による各種の個人用システムやエージェントが実現されることが見込まれる。

（6）自然言語処理の発展動向

自然言語処理は、自然言語文や文書の解析と理解を対象とし、機械翻訳、情報抽出、質問応答など多くの応用分野を射程とする人工知能の中核的な分野である。文の単語への分割や品詞付与、統語解析などの文構造の解析、単語や文の意味表現や意味的類似性、照応解析や文章構造解析など文脈処理などの言語解析技術を基盤としている。あらゆる解析過程で曖昧性があり、曖昧性解消のための言語知識に関する研究やタグ付きデータへの機械学習技術の適用により、解析精度の向上が図られている。ツイッター、ブログ、評価サイトなど消費者自らがコンテンツを生成するメディアである

consumer generated media からニュース記事や科学技術文書など様々な言語データからの情報抽出が課題となっている。以下では、「自然言語解析技術の高度化」、「基盤技術と言語資源の高度化」、「適用分野や対象の拡大」、という3つの観点から、自然言語処理の発展動向を、現在の動向、5年後の動向、および、10年後の動向について概要をまとめる。

現在の動向

「自然言語解析技術の高度化」に関しては、単語分割、品詞付与、統語解析など個々の文を対象にする言語解析技術は、ある程度の規模（100万語程度）のタグ付きコーパス（新聞記事、小説、webなどの言語テキストに品詞や統語構造などが付与された体系的なデータ）と大規模なタグなしコーパス（情報の付与が行われていないテキストデータ）の利用により、かなりの解析精度を達成することが可能になっている。また、このようなコーパスが豊富に収集されたリソースリッチな言語から、十分なコーパスが収集されていないリソースプアな言語へタグ付き情報を自動的に写像する手法などが研究されている。述語項構造解析*（動作や状態を表す語（述語）とその主語や目的語など（項と呼ばれる）の関係の解析）などの意味解析、照応解析や共参照解析などの文脈解析も盛んに研究されているが、まだ十分な精度が得られていない。「基盤技術と言語資源の高度化」に関しては、様々な機械学習手法、特に深層学習に基づく手法が広く適用されて、多くのタスクで高い性能を達成するようになった。個々の単語をベクトルとして空間に埋め込む手法である Word2vec*などの単語の表現学習や、sequence-to-sequence ニューラルネットなど構造を出力することができる深層学習法が牽引的な役割を演じるようになった。「適用分野や対象の拡大」に関しては、従来の新聞記事やwebテキストに加えて、様々な consumer generated media、社内文書、科学技術文書、特許文書など解析対象の文書が拡大している。質問応答、機械翻訳、対話など文や発話を生成する領域への応用が進展している。

5年後の動向

「自然言語解析技術の高度化」に関しては、句や文レベルの意味表現や表現学習の技術が高性能化すると考えられる。また、述語項構造などの意味解析や共参照解析などの文脈解析技術も高精度化する。それに伴い、文の意味的類似性や含意関係など文間の関係解析の技術が進展する。「基盤技術と言語資源の高度化」に関しては、深層学習技術がコモディティ化し、構造学習の手法が大きく進展すると考えられる。また、単語や文の表現学習と言語制約との融合技術が進展する。文の類似性や含意関係の解析に必要な表現技術やリソースが充実すると考えられる。「適用分野や対象の拡大」に関しては、人間と協調し、説明を必要とする自然言語応用、例えば、翻訳支援、読解支援、言語学習支援など、応用分野が拡大する。

10 年後の動向

「自然言語解析技術の高度化」に関しては、文脈を考慮した自然言語解析技術が進展し、文脈の意味を表現する技術の研究が進むと考えられる。特に、短い文脈や短い文書の文脈構造や意味を表現する方法の研究が進む。「基盤技術と言語資源の高度化」に関しては、大規模なタグ付きデータを用意することが困難なタスクに対して、半教師つき学習や学習データの半自動生成など、学習データを拡張する技術が進展すると考えられる。「適用分野や対象の拡大」に関しては、マルチモーダルな環境での自然言語処理応用へ、応用対象が広がる。言語情報以外の文脈も考慮した機械翻訳、質問応答、対話システムなどの応用システムが構築される。さらに利用者の目的に応じた文書要約や読解支援など、人とのインタラクションを伴う言語処理応用が進展する。

(7) ロボットの発展動向

ロボットは、自律的もしくは半自律的に作業を行う機械と定義できる。作業の目的は、産業（製造、農林水産、運輸など）、医療、エンターテインメント、軍事、宇宙、と多岐に亘る。AI・ビッグデータの利活用は、現時点でのロボット技術の大きな弱点である汎化・学習性能の低さを革新的に向上させるブレークスルーとして期待されている。機械学習の短所として途中過程のブラックボックス性が挙げられるが、行うべき動作など最終判断さえ得られれば良いことの多いロボット分野は、格好の応用先であるともいえる。以下では、「汎化能力の向上」と「学習能力の向上」の二つの観点から発展動向を考察する。

現在の動向

「汎化能力の向上」に関しては、対象物の特徴ベクトルとその対象物に対して取るべき動作を、機械学習で直接的に結び付ける End-to-End*の手法が展開されている。例として、ある物体を把持して持ち上げるとき、掴みに行く物体が何であるかを必ずしも認識する必要はなく、あくまでも（通常2指の単純構造の）ロボットハンドが安定に掴みに行けるハンドの位置・姿勢が得られればよいという考え方に基づき、RGB-D（カラー(RGB)と距離(Depth))画像上の入力パターンから直接、把持するためのハンドの位置と姿勢を深層学習によって求める手法が提案されている[3]。280種の物体の1000枚近い観測データ画像を学習に用いて、90%を超える検出成功率を得ている。対象物ごとの幾何モデルが不要なく、画期的に多種多様な対象物が扱えるようになり、工場など限られた環境でない、日常生活でも役に立つロボットの可能性が見えてきた。

「学習能力の向上」に関しては、高い認識性能を得るには、より多くの学習データを用いることが一つの重要な要素であるが、多数の学習データを獲得するのに多大な時間が必要となる。これに対して、2016年、Googleでは、複数のパターンを異

なる（14 台の）ロボットで、同時並行、分散型に学習し、日々、これらのデータを集結し学習した結果を再度分散し、新たな学習を行うことにより、人間よりも高速に、うまくつかむ工夫を学習してみせた[4]。機械の特性である「全く同じシステムで並行に学習し、その頭脳を完全に共有できる」を活かせば、ネックと思われた大量学習の負荷も問題ではなく、人を凌駕できる可能性を実際に示したものであり、インパクトが大きい成果である。

5 年後の動向

「汎化能力の向上」としては、クラウド上のロボット作業用のデータベースの構築が進み、これまでの各個別状況での学習から、大容量の共有データを活かした学習を行えるように遷移し、組み立て作業のようなより複雑で高度な作業にまで、人工知能技術と融合したロボット技術が展開される。例えば、ねじ回しを見つけて持ち上げ、ねじを締めるなど道具の使用がより汎用的に行われることが期待される。

「学習能力の向上」に直結する多数の学習データ獲得に関しては、物理シミュレーションで得られたデータとロボットが実際に動作して集めたデータを融合して活用する研究が進む。精度は落ちるが大量のシミュレーションデータと少量の高精度実データとの組み合わせの他、シミュレータで共通的な識別器を作成し、各ロボット個体による実データで転化学習するような枠組みも生まれる。

10 年後の動向

「汎化能力の向上」に関しては、異なる状況で獲得された知識のインテグレーションや更新が大規模に行われるようになり、未知の状況への適応能力が高まる。「学習能力の向上」に関しては、ロボット特有の能動性を活かし、学習の過程で十分でない情報を能動的に獲得する機能の研究が進み、自律性が高まる。

（８）マニファクチャリングの発展動向

マニファクチャリング*では、製品の開発・設計、製造準備、製造実行に関わる情報処理技術およびコンピュータシステムを対象としている。現状で代表的なものとしては、製品モデリング、製品モデルをコンピュータの内部に作成することを支援する CAD (computer-aided design)、CAD との併用で製品挙動などのシミュレーションや解析などの工学的な検討を行うことを支援する CAE、製品モデルに基づいて工作機械やロボットなどの制御データの生成を支援する CAM、製造に関わる全プロセスおよび製品ライフサイクル全体における技術的検討・評価を行うデジタルエンジニアリング、さらに最近では、CPPS (Cyber Physical Production System) が注目を集めている。以下、「バーチャルマニファクチャリング」、「製造装置の自律化・知能化」、「インテリジェント製品とその開発支援」の 3 つの観点から今後の動向について述べる。「バ

「バーチャルマニュファクチャリング」とは、デジタルファクトリ（仮想工場）での製造シミュレーションにより生産性や環境影響など最適な工程計画・作業計画を作りその結果を利用して実ラインを運用する技術である。「製造装置の自律化・知能化」は、工作機械、産業ロボットなどの製造装置が、製造対象に対して自動的かつ自律的に最適な動作で対応したり、装置自身が予防保全や他装置との協調動作を自律的行なう技術である。「インテリジェント製品とその開発支援」は、より快適で安心安全な日常生活を支える製品の出現とその製品の開発・設計と製造準備を支援する新しい CAD/CAE や CAM システムなどの開発のことである。

現在の動向

「バーチャルマニュファクチャリング」では、機能や振舞を含むモデリング技術、マルチエージェントシステム化技術、製造知識のデジタル化技術の開発が進められつつあり、シミュレーション精度向上や適用範囲拡大が進んでいる。また、モデリングやデジタル化では、スマートマニュファクチャリングを支える技術として新しい国際標準設定の努力が開始されている。「製造装置の自律化・知能化」では、主にプログラムにより動作が決まる装置について、これまでの作業データを利用した機械学習や深層学習、パターン認識技術などにより、より高度な最適化、自動化、自律化を実現する技術の開発が進められつつある。「インテリジェント製品とその開発支援」では、自動運転車、介護ロボット、デジタル家電など、いわゆる人工知能搭載の日常生活製品の開発が始まっている。一方、CAD/CAE/CAM など製品の開発支援技術は従来の考え方のままで構成している。

5年後の動向

「バーチャルマニュファクチャリング」では、装置エージェントが連携してマルチエージェントシステムとして仮想生産ラインを構成する技術の確立や、実運転データとシミュレーションデータを利用した生産システムの最適運転制御技術などの開発が進むと予想される。また、これらの実現に必要な国際標準もかなり整うと予想できる。「製造装置の自律化・知能化」では、装置自身で状態データを収集し、そのデータと与えられる製造対象物のデータを統合利用して、自身の判断でより最適な装置の運転をする技術の開発などにより、個々の装置の自律化・知能化が進む。「インテリジェント製品とその開発支援」では、機械学習、パターン認識など人工知能技術をソフトウェアとして一般の製品に組み込む技術の確立により、顧客の個別希望に合わせたカスタマイズ製品の実現が可能となる。

10年後の動向

「バーチャルマニュファクチャリング」では、実績データに基づいて、製品に合わせて最適な製造を実施する CPPS（Cyber-Physical Production System）ベースの生産

システム構成が定着していると考えられる。これにより、国際標準も整い、工場同士が連携し、広く製造知識も含む情報共有を行なう仕組みもできあがる。「製造装置の自律化・知能化」では、さらに知能化が進み、製造装置が自身の機能・振舞いも含んだモデルを自身の中に持ち、Cyber-Physical 一体装置とする技術が開発され、装置自体が CPPS の主要な構成要素なることも予想される。「インテリジェント製品とその開発支援」では、顧客が、欲しいもののイメージを入力すると、世界中の製品情報を参照して自動設計する技術や、CPPS と連携し製造できる工場へ繋ぐ技術が開発され、製品のオーダーメイド支援環境の整備が期待できる。

(9) ネットワーキングの発展動向

ネットワーキングとは、知識や情報を共有することにより人間社会を豊かにするための技術である。その技術を生かすためには、人間社会がそれに対応して変わることが必要である。発展動向の観点を、「基盤技術」（情報共有、同期／非同期コミュニケーション、共有仮想環境、ナレッジマネジメント、ゲートウェイインタフェース、セキュリティなど）、「適用範囲」（企業活動、教育システム、社会活動など）、「社会の成熟度」（情報リテラシーやコンピュータリテラシーの社会への浸透や、人間社会の知的な活動の度合い）とする。

現在の動向

「基盤技術」については、クラウドコンピューティング、ユビキタスコンピューティング、ソーシャルネットワーキングがすでに実用段階に入っており、さまざまなグループウェアも開発され、使われるようになっている。「適用範囲」については、グループウェアは、企業において、会議支援や協調作業支援などとして、導入が進んでおり、在宅勤務が例外的取り扱いではない企業も多くなっている。企業では教育・学習支援としての導入も始まっているが、大学においては、まだ遅れている。電子商取引（e コマース）は、すでに、一般社会にかなり浸透している。IoT（センサ情報）や人工知能（ボットやエージェント）との融合は始まっているが、まだ、第4次産業革命は本格化していない。「社会の成熟度」については、利用者の無防備な使用と、悪意をもつ者の存在が、社会不安を引き起こしているように、SNS を安易に利用することによる弊害が目立つようになっている。コピーアンドペーストの蔓延も社会問題となっているように、人間社会における倫理観の欠如も表面化しつつある。個人情報の取り扱いについては過渡期であり、個人情報の漏洩は問題だが、それを防止しようとして過剰に保護する傾向がある。

5 年度の動向

「基盤技術」については、携帯端末（スマートフォン等）やクラウドを活用したコ

ラボレーションシステムがさらに進展している。サイバーセキュリティ対策が進んでいる。「適用範囲」については、分散型企業活動は、技術の進歩に応じて、さらに進むなど、第4次産業革命が急速に浸透し始めている。遠隔教育が実現し、それにより、教育制度が変革する。たとえば、優れた内容の授業のネット配信が進み、レベルの異なる何種類かの授業内容が作成されるようになる。また、教員と学生が選択し、適切なカリキュラムを構築することができるようになり、授業にはそれを活用することにより、教員は、個別指導が必要な場面だけに注力することができるようになる。IT技術を活用して、研究機関における研究データ等の蓄積と活用が適切に行われるようになり、オープンサイエンスが進展する。一般社会においては、医療、福祉、防災、行政等の場面におけるネットワークサービスが進む。また、就職活動においても、求職者と求人者のマッチングの高機能化が進む。「社会の成熟度」については、初等中等高等教育の場において、倫理教育（研究倫理、技術倫理、生命倫理、等々）が浸透し、個人情報適切に保護され、かつ、必要な情報が共有できる社会が実現する。

10年後の動向

「基盤技術」としては、全く新しい情報通信技術が出現する。それにより、新たなロードマップが描かれるようになる。「適用範囲」はさらに拡大し、自分の分身を作ることができるようになり、それを使いこなすことができるようになる。「社会の成熟度」も高まり、それにより、社会の知的化が進む。

(10) フィンテック・電子契約の発展動向

フィンテック*とは、“finance”と“technology”を組み合わせた造語であり、2010年以降、IT技術を応用した様々な新しいサービスが提供されつつある。対象は、「決済」、「融資」、「海外送金」、「PFM（個人向け資産管理）」、「クラウド会計」、「投資」、などの金融関係のサービスから、さらには、執行機能を持つ契約書のIT化や組織の形成を自動化するDAO（分散型自動化組織）など、社会における商取引や契約全般に及ぶ。ITによって既存サービスの効率化を促進するだけでなく、リアルタイム的に顧客と一体になり新たな付加価値を生成する仕組みも内蔵できるため、金融機関のみならず、企業のあり方を大きく変貌させる潜在力がある。

以下では、「会計業務ソフト」、「ブロックチェーン技術」、「サービスデザイン」という3つの観点から、フィンテック・電子契約*の発展動向を、現在の動向、5年後の動向、および、10年後の動向について概要をまとめる。

現在の動向

「会計業務ソフト」に関しては、個人向け・企業向けから国をまたがる国際会計まで、お金の出入りに関する様々な複雑な処理を自動的に行う技術が次々と開発されている。「ブロックチェーン技術」は、ビットコインが実用化された代表であるが、分散

型のコンピュータネットワークにデータを共有することで、廉価かつ高速に信頼性の高い取引を可能とする技術として、様々な企業が実証実験を行うなど、潜在的な可能性を探っている。「サービスデザイン」に関しては、顧客ごとのリスク嗜好に応じた資産運用を提案できるようなロボアドバイザーが実用化し、フィナンシャルプランナーの業務を代用できるようになってきている。

これらの技術の適用対象は、「決済」、「融資」、「海外送金」、「PFM（個人向け資産管理）」、「クラウド会計」、「投資」、などの金融関係のサービスから、さらには、執行機能を持つ契約書のIT化や組織の形成を自動化するDAO（分散型自動化組織）など、社会における商取引や契約全般に及ぶ。ITによって既存サービスの効率化を促進するだけでなく、リアルタイム型に顧客と一体になり新たな付加価値を生成する仕組みも内蔵できるため、金融機関のみならず、企業のあり方を大きく変貌させる潜在力がある。

5年後の動向

「会計業務ソフト」に関しては、多くの企業が使用するようになる。オープンAPIが進み、さらに高度化したサービスの提供が可能となる。商流情報も活用されるようになり、データに基づく融資先の信用リスク評価も可能となる。「ブロックチェーン技術」に関しては、現状では高額の手数料がかかる海外送金が廉価に高速に処理できるようになる。現実の通貨と同等の信用性を持つような仮想通貨が流通するようになる可能性がある。「サービスデザイン」に関しては、また、クラウドファンディングが広まり、一般の人が少額からの資金運用を手軽に行えるようになる可能性がある。

10年後の動向

「会計業務ソフト」に関しては、世界的にクラウド会計が個人でも企業でも標準的になり、クラウド内に蓄積されたビッグデータを活用した新しいサービスが次々と提供される。海外など遠隔地と高い信頼性を伴ってビジネス契約書を交わすようなことも、短時間で実行できるようになり、企業間の取引が世界的に大きく広がる。「サービスデザイン」に関しては、分散型自動化組織が実用化され、目的と狙いに賛同する人を集めて電子契約を交わして組織を稼働するような新しいタイプの企業が誕生する可能性がある。

(11) 行政ミクロ情報等の利活用の発展動向

欧米並びにわが国では公的統計作成のために収集された統計ミクロデータや行政情報の研究利活用がデータ駆動立国を支える基盤の方策と考えられている。既に欧米では、統計ミクロデータの研究利活用基盤としてのオンサイト施設が設置され、証拠に基づく政策立案(EBPM、Evidence Based Policy Making)等に関わる研究利用がある程度可能になっている。わが国が目指すオンサイト施設は、中央データ管理施設と回

線をつなぎ、シンクライアント環境で統計マイクロデータ等をダウンロードすることなく探索的データ分析を可能とする拠点である。以下では、「証拠に基づく政策立案」、「オンサイト施設」、「データ融合」の観点から、現状、5年後、10年後の動向を考える。

現在の動向

「証拠に基づく政策立案」に関しては、わが国では、公的統計マイクロデータ並びにそれに関連する行政情報のオープン化をある程度可能とする法制整備が2018年以降に実施される[5]。これに伴い「オンサイト施設」については、2017年1月から総務省統計マイクロデータによるオンサイト施設試行運用が、神戸大、一橋大、滋賀大、情報・システム研究機構と(独)統計センターとの間で開始された。そこでは、オンサイト施設で行ったデータ分析結果の持ち出しに関する審査を行うデータホルダーとしての総務省、中央管理施設としての(独)統計センターによる審査体制が検討されている。2018年には、和歌山に設置される「総務省統計局統計利活用センター」を含む約10施設による本格運用が開始される。「データ融合」に関しては、新しい家計インデックス(CTI)の2018年提供に伴い、総務省統計局は民間ビッグデータと統計マイクロデータ情報とのデータ融合を支える産官学連携組織の形成を開始する[6]。これは状態空間モデルやPropensity Score*を用いた、統計マイクロデータと民間ビッグデータとのデータ融合活動の嚆矢となり、官民への波及が期待される。

5年後の動向

「証拠に基づく政策立案」に関しては、各府省には2018年以降、EBPMをマネジメントするEBPM推進統括官が設置されるとともに、政府横断的な取り組みも官民データ利活用戦略会議の下に置かれるEBPM推進委員会が担当する[7]。その中で、行政マイクロ情報はEBPMの基盤として位置づけられたために、これを分析し政策立案・効果予測・評価などを実効化する必要性が高まる。

「オンサイト施設」については、今後5年内外で、総務省以外の統計マイクロデータと一部関連行政情報を分析可能なオンサイト施設を全府省、地方自治体、全国各県の研究機関に設置することで、EBPMとそれを支援する公共性の高い学術研究の加速が予期されている。

「データ融合」に関しては、全府省・自治体・全国学術機関において行政マイクロ情報と民間ビッグデータとを融合して分析可能なオンサイト施設の構築、総務省統計研究研修所による行政データサイエンティストの育成が進められる。

10年後の動向

「証拠に基づく政策立案」に関しては、行政マイクロ情報を匿名化したPublic Use Dataの提供、民間保有ビッグデータと行政情報との融合を容易にする変換テーブル

の提供など、民間も行政情報を利活用できる仕組みの検討も進捗し、官民一体のデータ利活用推進の取り組みを通じて政策の質向上のみならず、データ利用環境整備を通じた国際産業競争力強化にも資することが期待される。

また「オンサイト施設」については、秘密計算処理などを活用して、マイクロ情報閲覧不可能なデータ分析環境を開発し、倫理コードの順守を誓い、一定の力量を有するデータサイエンティストが自身の研究施設から行政マイクロ情報等を分析可能な新たなオンサイト環境の構築を目指すことで、データサイエンティストに支援された学術研究、政策提言などが大幅に増加することが期待される。

(12) 人工知能リスク管理の発展動向

人工知能（AI）リスク管理とは、AI 技術の適用に伴って好ましくない結果が発生するリスクに対して、事前にリスクを評価して適切な対策をするとともに、好ましくない事態の発生後に適切な対応をすることである。革新的な AI 技術が登場し不安も招いているなか、AI 技術が人間の心地よい社会につながるためには、リスクの抽出・評価、対策を継続的に行なうリスクマネジメントに、研究開発と並行して留意すること必要である。ここでは、AI の適用分野には依らないで、専ら機械学習が誤るリスク等への「技術的対策の高度化」と、故意または意図せずにもたらされる全般的なリスクへの「社会的対策の高度化」との二つの観点で、リスク対策の動向を概観する。

現在の動向

「技術的対策の高度化」に関しては、深層学習による AI は、仕様に基づいて開発された IT システムのように品質を保証することが困難である。利用者の意思決定支援に利用される場合、AI の出力を人が判断して利用できるよう、説明機能など人が AI の能力を理解するための技術が試みられている。

「社会的対策の高度化」に関しては、AI への期待が膨らみ、制御不能に陥る、人の職が奪われる等の不安が議論されており、開発者視点中心に、AI 開発原則や倫理指針などのガイドラインが策定されつつある。

5年後の動向

「技術的対策の高度化」に関しては、AI の出力が自動的に機械の作動や注文の入力につながる局面が増えると、AI を採用するか否かを判断できるよう、利用者が AI を評価・選択するための技術が望まれる。

「社会的対策の高度化」に関しては、AI の利用が進展すると、期待から反転して AI への失望が見られ、理解不能、隠匿された悪意などの具体的な不安が生じるであろう。AI によるサービスの提供者視点で、AI の責任範囲の明確化や、事故・犯罪情

報とベストプラクティスの共有、事故・犯罪から人を守る警察的な機能等も期待される。

10年後の動向

「技術的対策の高度化」に関しては、AI の出力が自動的に他の AI の入力となる AI 同士が連携する局面が増えると、ある AI が誤っても、部分停止や修復するようなロバストな AI のネットワークを実現する技術が、望まれる。

「社会的対策の高度化」に関しては、AI が人と共存することが進むと、AI への不応感を抱く人への対応も課題となるであろう。利用者の視点で、教育、支援、監視、緊急対応、保険などの機能や職種転換支援などのセーフティネットの構築も期待される。

<用語解説>

スパースモデリング

超高次元の一見複雑なデータでもその中に含まれる本質的情報は比較的低次元の構造を持つというスパース性（疎性）を想定して、情報を効率的に抽出する方法の総称。超高次元データを大量に観測すること、あるいは解析することが非現実的な場合でも、スパースモデリングによって、少ないデータ量や現実的な計算量で、必要な情報を的確に取り出すことができるようになる。

ムーアの法則

大規模集積回路の集積度は18か月で約2倍になるという経験則で、インテルの創業者ゴードン・ムーアの1965年の論文にちなんで名づけられた。これに従うとすると集積度は10年後には約100倍、20年後には約1万倍となる。ムーアの法則は、むしろ製品開発の目標値としての役割を果たし、50年後の現在に至るまで、記憶容量や演算速度はほぼこの経験則に沿った形で発展してきた。

エッジ（コンピューティング）

データの蓄積と処理をすべてクラウドに集約するのではなく、データが生み出される場所（エッジ側）の近くの処理基盤に一部を任せることです。その結果、クラウドサーバ側の負担を軽減し、全体としてデータの処理の遅延を極小化し、リアルタイム性を維持することを可能にします。

flops/s 指向

コンピュータの処理能力として、理論演算性能（Flop/s, floating point operations per second）の向上を重視して設計されたコンピュータのこと。

Bytes/s 指向

メモリやファイル I/O などデータ処理ユニットのデータ転送レート（Bytes/s）の向上を重視して設計されたコンピュータのこと。

データスループット指向

コンピュータの処理能力として、単位時間あたりどれだけ多くのデータを処理できるか（スループット）に重点を置いて設計されたコンピュータのこと。

More-Than-Moore 型, Moore 型, Post-Moore 型

Moore 型とは Moore の法則による半導体技術の微細化の恩恵をうけて発展進化してきた現在のコンピュータの総称、More-Than-Moore 型とは、Moore 型の集積回路技術の

微細化だけではなく、機能デバイス等より高度なデバイス（センサーや通信モジュール等）をシリコンチップ上に集積する技術で作られたコンピュータ、Post-Moore 型とは、量子コンピュータなど、トランジスタで構成される現在のコンピュータとは異なる動作原理を用いて計算処理を行うコンピュータのこと。

情報トリージ技術

限られた時間的制約の下で膨大かつ玉石混交な情報の中から意思決定や問題解決に有益な情報を効果的に峻別・整理する方法。

インメモリコンピューティングアーキテクチャ

ソフトウェアを実行する際、使用するプログラムやデータのすべてをメモリ上に読み込み、ハードディスクなどの外部記憶装置を使わないで処理を進めるコンピュータのこと。

サイバーフィジカルシステム

サイバーフィジカルシステム（CPS） CPS とは、実世界（フィジカル空間）にある機器やインフラなどに関する多様なデータをセンサーネットワーク等で収集し、サイバー空間で大規模データ処理技術等を駆使して分析／知識化を行い、そこで生み出された情報によって、それらの障害を未然に防いだり最適制御をおこなうことを可能にするもの。

Spark

Hadoop に続くビッグデータ分散処理のための新しいフレームワーク。

データリネージ

データの有効活用のために、データがどのように抽出、変換されているかを追跡し、それを記録していくこと。

データ同化

偏微分方程式を数值的に解く計算機シミュレーションの結果を、実際の観測・計測データとなるべく合うように、シミュレーション内の初期条件、境界条件、パラメータ等を調整する計算手法。

生体に非侵襲の計測装置

生体内の計測対象および関連部位に直接触れることなく、電磁波や音波を用いて極力生体に影響が残らない形で計測を実現する装置。身近なものとして、CT（コンピュータトモグラフィ）がある。

マルチモーダルデータ

視覚や聴覚などの複数の感覚に関わるセンサーを用いて計測されるデータ。

各種エージェントの方策

各種エージェントがその環境で選択する行動についての報酬が最も多く得られるような方策。

述語項構造解析

自然言語の単文は、動詞・形容詞・動作名詞など動作や状態を表す語（述語）と、その主語や目的語など主要な関連概念（述語に対する項）よりなる。文中のそれぞれの述語に対し、その意味上の主語や目的語などの項を同定する技術を述語項構造解析という。

Word2vec

単語の意味を表現するため、類似の文脈に現れる語は類似の意味をもつといういわゆる分布仮説に基づき、同じ単語の近辺に現れる語がベクトル空間上の近隣の位置に泡われるよう各単語に対応するベクトル表現を繰り返し学習する手法。

End-to-End

前処理や途中過程の分解など行うことなく、直接、入力情報から出力情報を導くモデル（手法）を学習するアプローチ。

マニファクチャリング

生産・製造など「ものづくり」関連の情報処理技術。主に、工業製品の設計・製造・利用・保守・リサイクルなどを支援する情報処理技術。

フィンテック (fintech)

“金融(finance)”と“技術(technology)”を組み合わせた造語であり、従来よりも、安く、速く、便利に金融商品やサービスを提供する新しい IT 技術全般を意味する。2010 年ごろから金融機関のみならず中小企業や個人企業にまで急速に普及しつつある。似た言葉である“フィナンシャルテクノロジー (financial technology)”は、金融工学と訳され、金融商品の価格変動を分析し応用する工学的研究であり、意味が異なる。

電子契約

従来の書面による契約に代わり、電子ファイルを利用した契約であり、印紙代や印刷費、保管のための経費などが不要となり、電子メールやホームページ上で速やかに契約を結ぶことができる。日本では、2001年に施行された電子署名法により、公開鍵暗号方式に基づく電子署名が署名や押印と同等の法的効力を持つことが定められ、多くの企業で使われるようになってきている。

Propensity Score

母集団から、あるデータが抽出され観測される確率の推定値の逆数。このスコアを抽出された標本データのある特性の加重平均の重みとして利用することで、偏った標本抽出であっても母集団の特性を偏り無く推定することが出来る。この確率の推定には、データが抽出されるか否かといった情報とそれに影響を与える補助情報との両者が母集団全体で観測されている必要がある。

<参考文献>

- [1] 河原達也, “音声認識技術の現状と将来展望”, 電気学会誌, 133(6), pp. 364-367, 2013.
- [2] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks”, Advances in Neural Information Processing Systems 25 (NIPS2012), 2012.
- [3] I. Lenz, H. Lee, A. Saxena, “Deep Learning for Detecting Robotic Grasps”, The International Journal of Robotics Research 34 (4-5), pp.705-724, 2015.
- [4] Google Research Blog: <https://research.googleblog.com/2016/03/deep-learning-for-robots-learning-from.html>
- [5] 統計改革推進会議: 3. ユーザの視点にたった統計システムの再構築と利活用促進, 「統計改革推進会議最終とりまとめ」, pp. 13-20, 2017/05/19, 首相官邸: http://www.kantei.go.jp/jp/singi/toukeikaikaku/pdf/saishu_honbun.pdf
- [6] 速報性のある包括的な消費関連指標の在り方に関する研究会: 消費動向指数(CTI)の開発に向けて, 2017/03/22, 総務省: http://www.soumu.go.jp/main_content/000473104.pdf
- [7] 統計改革推進会議: 1. EBPM 推進体制の構築, 「統計改革推進会議最終とりまとめ」, pp. 1-6, 2017/05/19, 首相官邸: http://www.kantei.go.jp/jp/singi/toukeikaikaku/pdf/saishu_honbun.pdf

＜参考資料＞

情報学委員会 e-サイエンス・データ中心科学分科会 審議経過（開催日時と主要議題）

第1回 平成27年3月18日

- ・委員長、副委員長、幹事の互選
- ・前期の活動報告
- ・各委員からの報告
- ・今後の活動方針について

第2回 平成27年9月28日

- ・マスタープラン2017について
- ・講演（新井委員）
- ・今後の活動について

第3回 平成28年1月12日

- ・講演（有田委員）
- ・今後の活動について

第4回 平成28年5月12日

- ・講演（鷺尾委員）
- ・講演（中島委員）
- ・今後の活動について

第5回 平成28年10月3日

- ・講演（上田委員）
- ・「AI・ビッグデータ利活用のトレンドとインパクト」のアンケート調査結果について

第6回 平成29年1月12日

- ・ビッグデータ利活用のトレンドとインパクト
ー調査報告のまとめと意見交換ー

第7回 平成29年6月8日

- ・AI・ビッグデータ利活用展墓調査報告書
- ・その他

掲示板による審議（平成29年7月6日ー13日）

第8回 平成29年7月13日ー18日（メール審議）

- ・記録（案）の承認

公開シンポジウム

- ・「第19回大阪大学産業科学研究所国際シンポジウム」平成27年12月7ー9日（大阪大学，銀杏会館）
- ・「by 機械学習、of 機械学習」平成27年11月24日（日本学術会議，講堂）

計算基盤の発展動向

計算基盤とは、超並列処理など最新の計算技術を用いて大規模シミュレーションや大規模データ処理を行うための基盤となる高性能計算機システムのことである。

発展動向の整理軸	現状の動向	5年後の動向	10年後の動向
性能向上	例：計算科学分野でのHPC(高性能計算)技術の発展を受けて、超並列処理やベクトル処理を大規模データ解析に活用	現在の計算性能重視(flop/s指向)のアーキテクチャ(Compute-Centric Architecture)設計から、大規模データ解析に必要なメモリスループット(B/s指向)や大規模メモリ重視のメモリ中心アーキテクチャ(Memory-Centric Architecture)設計への変換	3次元実装・シリコンフォトニクス等More-Than-Moore技術を活用して、レジスタ・キャッシュ・メインメモリ・ファイルシステムを統合的に扱うユニバーサルメモリアーキテクチャとその各階層のインテリジェント化のためのインメモリコンピューティングアーキテクチャの確立
AI・ビックデータ向けアーキテクチャの高度化	次々に生み出される様々なAI・ビックデータ利活用のためのアルゴリズムの汎用コンピュータへの実装	AI・ビックデータ利活用において、ある程度応用分野を限定し、それぞれの分野に適した共通プラットフォームとしての計算モデルの構築とそのコンピュータ上での仮想化による実現	仮想マシンを実現する「マイクロアーキテクチャ」設計に、マルチコア・メニーコア、FPGA、精度保証可変長計算、ユニバーサルメモリ、インメモリコンピューティングなど様々な従来のMoore型マイクロアーキテクチャとPost-Moore型アーキテクチャとの融合
計算基盤構築技術	IoTセンサー群をスタンドアロンコンピュータ・サーバに接続し、集中処理	IoTのインテリジェント化(AIoT化)の発展により、エッジでのデータのフィルタリングとメタデータ化(情報トリアージ技術)、サーバ側での高次データ解析技術とそのAIデータベース構築技術	仮想マシン上に構築されるデAIデータベースのオープンデータ化とインタークラウド技術によるAI連携(分散共有処理)の発展

ビッグデータ，データベースの発展動向

ビッグデータとは、既存の一般的な技術では管理することが困難な大量のデータのことである。より広義には、多様性、データ量、データの生成頻度、更新頻度の点で管理が困難なデータ、および、それらを蓄積・処理・分析するための技術を含む概念である。データベースとは、検索や蓄積が容易にできるよう整理されたデータの集まりのことである。データベースへのデータの格納方法を論理的・物理的に規定するデータモデルには、階層型データモデル、リレーショナルデータモデル、オブジェクトデータモデルなど、いくつかの代表的モデルが知られている。

発展動向の整理軸	現状の動向	5年後の動向	10年後の動向
アルゴリズムの高速化 ビッグデータを効率よく扱うための新しいアルゴリズムパラダイムを構築し、ための革新的なアルゴリズム技術の開発	スーパーコンピュータを用いて数十億点のグラフに対する探索や最短路計算が高速にできる	多くの問題に対して、データサイズに依存せず、近似的に定数時間で、解を生成する乱択アルゴリズムが実用化されている	超大規模なグラフに対する最短路計算などのさまざまなアルゴリズムを通常のパソコン上で近似的に解く方法が実用化され、諸分野へ応用されている
ビッグデータに対するデータベース技術の革新 ビッグデータに対処できるような分散処理環境の構築、IoTからのセンサーストリームデータの蓄積と解析をサポートするようなデータベース技術。	ビッグデータを大規模分散環境(e.g., map reduce)を用いて数時間～over night程度の計算時間でデータ解析可能になった。オンメモリ分散処理(e.g. spark)による準リアルタイム処理も視野に入っている。	IoTからのセンサーストリームと蓄積されたビッグデータとセンサーデータが融合し、機械学習・ビッグデータ解析が可能になる。データのクオリティ(精度、信頼度、フォーマット、リネージ、プライバシー)をコントロールするための技術が実用化され始める。	蓄積データやセンサーデータを覆う統一的な情報空間の上で、クオリティデータのクオリティ(精度、信頼度、フォーマット、リネージ、プライバシー)がコントロールされた異種データが、解析者の必要に応じて高い自由度で統合され、機械学習・ビッグデータ解析にシームレスに結合される。
リアルタイム処理技術の高度化 ストリームセンサーデータを圧縮しながら処理し、高速伝送する技術の総称	センサーデータなどリアルタイムで発生するビッグデータを圧縮しながら蓄積、処理するストリームアルゴリズムが実用化されつつある	高速伝送を実現するストリームアルゴリズムが、さまざまな機器に組み込まれ、交通、経営、医療の分野での実用化が実現している。	遠く離れた観測装置から得られる観測データを世界各地に高速伝送し、共有しながら解析ができるなど、科学分野での活用が行われている。

シミュレーション・データ同化の発展動向

シミュレーションは一般的に、基礎方程式(支配方程式)を用いて、時間発展あるいは空間変形を逐次的に大規模計算機(スーパーコンピュータ)を用いて数值的に解くことを言う。またデータ同化とは、このシミュレーション内の変数を、その一部を観測・計測したデータと突き合わせ適切に整合させる作業である。近年研究が生命や社会の複雑な現象の予測や定量的理解にまで拡大したことに伴い、基礎方程式を超高精度に解けない、あるいはそもそも基礎方程式が存在しない領域においても、微視的な半経験的モデルとマクロな観測・計測データを統合する、いわゆる情報統合の方策としてデータ同化の有用性は高まっている。

発展動向の整理軸	現状の動向	5年後の動向	10年後の動向
規模の拡大	シミュレーション単体では数千万から10億程度の変数を解くことが可能。データ同化に関しては、データの質および量にかなり依存するが、1000万程度の変数に対する同化がラフに可能。	10倍程度の変数を取り扱える。解像度で言えば、1/10程度。データ同化は対象の変動の性質にかなり依存するため、1000万程度の現状を少し上回る程度か。積分時間で言えば、50倍程度。	100倍程度の変数を取り扱える。積分時間では、1000倍程度。データ同化は1億程度の変数の同化が可能。
ソフトウェアのオープン化	空間メッシュの自動化はほぼ実現。鍊成シミュレーションも、ダウンスケーリングのような静的かつ準整合的方法は確立。データ同化のメタ化は未踏。	ダウンスケーリングの自己無撞着性(整合的解法性能)が向上。時空間発展方程式のデータ同化および、シミュレーション内のマクロパラメータの同化についてはメタ化が実現。	個人/グループレベルで作成するシミュレーションモデルの既存手法との連動技術が確立。メタデータとシミュレーションコードを入れると、適切なデータ同化が実現。
可視化・結果解釈機能の高度化	探索的に人が見たい変数(アスペクト)を選択し、視覚的に効果的に自動描画可能。アスペクトの自動選択機能が実現されはじめている。	アスペクトの自動選択が、スパースモデリングあるいは深層学習によって実現。その結果を利用したエミュレーションも実現。位相データ解析と連動した効果的描画機能の実現。	シミュレーションコードと結果の因果解析を相補的に利用することで、因果構造を半自動的に抽出する機能が実現。位相データ解析やエミュレーションも汎用化。

機械学習・深層学習の発展動向

機械学習とは、人工知能における研究課題の一つで、人間の学習能力と同様の機能を、明示的にプログラミングすることなくコンピュータで実現しようとする技術・手法である。機械学習アルゴリズムは、対象とするデータの生成機構を反映するデータ間の特徴的関係やパターンを抽出し、さらにそれらを用いて同じあるいは類似の機構から生成された新たなデータについて推定や予測を行う。

深層学習とは、多層のニューラルネットワークを用いた機械学習技術の総称であるが、特に層数が3を超える深いノード段数を有するニューラルネットワークを指すことが多い。従来のバックプロパゲーションに、スパース・コーディングやオートエンコーダなどの手法を組合せ、大規模なニューラルネットワークの学習を可能にした。

発展動向の整理軸	現状の動向	5年後の動向	10年後の動向
対象データの大規模化 事例数や属性次元数、関係の複雑さ・多様性など、様々な面で扱う対象データの規模	深層学習などで時空間的構造を持つ大量データからの高精度学習が実現された。ただし、膨大な学習データが必要。並行して少数データからの学習や意味理解可能な潜在的規則性発見の研究が進んでいる。	シミュレーションデータで仮学習して、実少数データに転移学習する方法が広く開発される。深層学習を組み込むシステム化技術により、大規模問題での学習実用化が加速する。大規模問題に対するリアルタイム逐次学習技術が進む。	時空間的構造を持つデータとして構造化出来ないデータも、高精度な学習が行える技術改良が進むか、あるいは一般的データを時空間的構造に埋め込んで高精度学習を行う技術が開発される。
学習の自律性 アクティブ事例選択、アクティブ属性選択、強化学習のように、対象データの内容や学習の目的に応じて、自律的に事例や属性を収集する能力と共に、自律的に学習目的や基準を選択、設定する能力	強化学習で、複雑な状態や文脈に基づいて、効果的な自律的情報収集・試行を行う学習技術が実現されている。深層学習に強化学習を導入し、大規模・複雑な学習を自律に行えるようになった。	左記の現状動向が深化する。	複雑・多様な問題構造を探索的に発見する研究が進み、限定的だが問題定義の置き換えを行い、フレーム問題を緩和する強化学習技術が実現する。学習目的・基準を自律判断するアクティブ学習が実現する。
対象問題の多様化 分類やクラスタリング、外れ値検出などの典型的な機械学習問題に加えて、情報検索やシミュレーション、計測、意思決定など他の情報処理問題を扱う理論や技術と融合による対象問題の多様化の度合い	外れ値が発生する以前の予兆発見・検知や因果推論、クラウド技術との融合研究が行われている。一部には、学習技術から新たな知見を創造する技術が発展しつつある。	より多くの他の情報処理問題を扱う技術と機械学習との融合が進み、より多様な問題を扱う機能を実現できるようになる。	異分野融合の大規模システムや複数分野のビッグデータを横断的に利用する機械学習技術・統計的アプローチが実現される。

パターン認識の発展動向

パターン認識とは、画像認識、物体認識、音声認識など、画像・音声などの雑多な情報を含むデータの中から、一定の規則や意味を持つ対象を選別して取り出す処理である。この処理では統計、機械学習技術が一般的に用いられ、その目的は識別や回帰に代表される。

発展動向の整理軸	現状の動向	5年後の動向	10年後の動向
・ 機能の高度化 機能の高度化とは、統計、機械学習技術の発展によるパターン認識の目的である識別や回帰の性能向上を指す。データに含まれる各種ノイズに対する頑健性の向上を含む。	深層学習によりパターン認識の性能向上がはかられている。音声認識では数千時間の学習データを用いた場合に10%以下の精度、一般物体認識では千カテゴリについて0.15%程度の精度を得ることができるようになっている。また物体のクラスや識別に有効な局所特報量が自動獲得されつつある。	深層学習の機械学習技術が深化し、人手による調整が必要であったネットワーク構造やパラメータの初期値の設定が自動化して、パターン認識の性能およびインタフェースが向上すると考える。同時にパターン認識の応用領域の拡大が見込まれる。一般物体認識においては、物体のクラス間の構造が自動獲得されることが見込まれる。	深層学習を超える統計、機械学習技術が開発され、パターン認識の性能およびインタフェースがさらに向上すると考える。いろいろな関係性(因果や階層構造等)が自動獲得されるようになると思う。また希少な事象のデータについても、高度なパターン認識と統計、機械学習を用いて解析できるようになると考える。さらに、自律的にデータ取得、学習を行い、パターン認識技術を自動で進化、再生する技術が研究開発されるようになると思う。
・ 機能の多様化 センシング技術の発展とともに、従来の音声や画像データのみならず、多様な種類のデータが観測され、利用できるようになっている。機能の多様化とは、それら多様化したデータのパターン認識の実現を指す。各種データのいろいろな組み合わせによるパターン認識の実現を含む。さらに従来の目的の識別や回帰を超えて、意思決定のための方策を自動獲得するなどの目的の多様化を含む。	例えば、スパムや攻撃メールの検出においてパターン認識が利用されている。また道路気象や交通規制の情報を含め複数のデータを利用した交通量管理システムが実現されつつある。さらに強化学習と組み合わせ、ロボット行動などの方策の自動獲得が実現されつつある。	例えば、一般物体認識について、マルチモーダルデータの自然な統合によるパターン認識が実現されることが見込まれる。またパターン認識を利用した自動運転技術が普及することが見込まれる。高度なパターン認識と強化学習を用いた、各種エージェントの方策の自動獲得が実現されることが見込まれる。	例えば、パターン認識による一般物体認識システムは、進化するIoT技術等を用いて、自律的に認識対象を広げていくと考える。
・ 機能の個性化 機能の個性化とは、各ユーザ(集団)の特性に対応したパターン認識の実現を指す。	ユーザの話し方に対応した音声認識、ユーザ環境を考慮した物体認識などの個性化技術が実現されつつある。	パターン認識の各処理のユーザ(集団)対応が実現される。	ユーザ(集団)対応したパターン認識による各種の個人用システムやエージェントが実現されることが見込まれる。

自然言語処理の発展動向

自然言語処理は、自然言語文や文書の解析と理解を対象とし、機械翻訳、情報抽出、質問応答など多くの応用分野を射程とする人工知能の中核的な分野である。文の単語への分割や品詞付与、統語解析などの文構造の解析、単語や文の意味表現や意味的類似性、照応解析や文章構造解析など文脈処理などの言語解析技術を基盤としている。あらゆる解析過程で曖昧性があり、曖昧性解消のための言語知識に関する研究やタグ付きデータへの機械学習技術の適用により、解析精度の向上が図られている。Twitter、blog、評価サイトなどのconsumer generated mediaからニュース記事や科学技術文書など様々な言語データからの情報抽出が課題となっている。

発展動向の整理軸	現状の動向	5年後の動向	10年後の動向
言語解析技術の高度化 単語セグメンテーション、品詞付与、チャンキング、統語解析、述語項構造解析、照応/共参照解析、意味表現、表現学習、語義曖昧性解消、談話解析、文書構造解析などの言語解析技術の高性能化	個々の文を対象とする言語解析では性能向上の余地はあるものの、高い性能を達成している。述語項構造など意味解析や文脈解析ではまだ実用レベルの解析精度に達していない	句や文の意味解析・意味表現・表現学習技術が高性能化。述語項構造解析や共参照解析など、文や文章の意味解析の基盤技術が高度化	文脈解析の高度化。意味表現技術、文脈表現技術の高度化。文の意味を文脈上で捉える技術や、短い文書中の文脈を表現する方法に関する研究が進展する。
基盤技術・言語資源の高度化 大規模言語データの処理基盤、機械学習の適用法、構造学習、同時学習、マルチタスク学習、表現学習、学習データの構築、半教師付学習、教師なし学習、能動学習などの基盤技術。レキシコン、コーパスなどの言語資源	様々な機械学習アルゴリズムがあらゆるタスクに適用されてきた。深層学習に基づく手法が広く適用され、多くのタスクで高い性能を達成している。表現学習やsequence-to-sequence modelが牽引的	DNN技術のコモディティ化。DNNに基づく構造学習の高度化。表現学習と言語制約の融合技術。同義・含意認識など文脈理解のための表現技術やリソースの充実	大規模なタグ付きデータが必要なタスクに対する半教師付学習やデータの半自動タグ付けなど、データ拡張に関する技術が進展
解析対象の拡大 機械翻訳、評判情報抽出、文書分類、質問応答、文書要約、情報抽出、関係・イベント抽出などの既存の応用から、文脈理解を必要とするマルチモーダル環境での言語理解、対話など対象の拡大	Consumer generated media、社内文書、科学技術文書、特許文書など対象の拡大。対話や質問応答などの発話生成などの応用への拡がり	他分野への応用。翻訳支援、読解支援、学習支援など結果の説明を要する新しい応用分野への適用	マルチモーダル環境での自然言語処理応用。文脈を考慮した機械翻訳、質問応答。対象者に応じた文書要約、読解支援。

ロボットの発展動向

ロボットとは、自律的もしくは半自律的に作業を行う機械。作業の目的は、産業（製造、農林水産、運輸など）、医療、エンターテインメント、軍事、宇宙、と多岐に亘る。AI・ビッグデータの利活用は、現時点でのロボット技術の大きな弱点である汎化・学習性能の低さを革新的に向上させるブレークスルーとして期待される。

発展動向の整理軸	現状の動向	5年後の動向	10年後の動向
汎化能力の高度化 既存のロボット技術は、各状況に応じ、特化したアルゴリズムを用いることで高性能を実現しているものが多く、状況が変わる（対象物が変わる、ロボットシステムの構成が変わる）と、新たなアルゴリズムの開発が必要であることが大きな問題である。人による個別状況への対応を必要としない汎用性の高い技術が望まれる。	深層学習を用い、対象物の特徴ベクトルとその対象物に対して取るべき動作を直接的に結び付けることにより、多種多様な対象物を同一に扱う枠組みが提案された。	クラウド上のロボット作業用DBの構築が進み、組み立て作業のような、より複雑で高度な作業まで汎用的に行われるようになる。	異なる状況で獲得された知識のインテグレーションや更新が大規模に行われるようになり、未知の状況への適応能力が高まる。
学習能力の高度化 異なる個体のロボット間で情報を共有することで、より豊富な知識、高度な技能を獲得する機構の高度化が望まれる。	複数の個体のロボットが並行して異なるパターンを分散的に学習し、その情報を交換、共有することにより、高速な学習が可能であることが示された。	物理シミュレーションで得た多数データとロボットが実際に動作して集めたデータの融合活用が進む。シミュレータで共通的な識別器を作成し、各ロボット個体による実データで転化学習する枠組みも生まれる。	ロボット特有の能動性を活かし、学習の過程で、十分でない情報を能動的に獲得する機能の研究が進み、自律性が高まる。

マニファクチャリングの発展動向

マニファクチャリングでは、製品の開発・設計、製造準備、製造実行に関わる情報処理技術およびコンピュータシステムを対象としている。現状で代表的なものとしては、製品モデリング、製品モデルをコンピュータの内部に作成することを支援するCAD、CADとの併用で製品挙動などのシミュレーションや解析などの工学的な検討を行うことを支援するCAE、製品モデルに基づいて工作機械やロボットなどの制御データの生成を支援するCAM、製造に関わる全プロセスにおける技術的検討・評価を行うデジタルエンジニアリングがある。

発展動向の整理軸	現状の動向	5年後の動向	10年後の動向
• バーチャルマニファクチャリング デジタルファクトリ(仮想工場)での製造シミュレーションにより生産性や環境影響など最適な工程計画・作業計画を作り、その結果を利用して実ラインを運用する技術。知識表現、プランニング、最適化、機械学習、マルチエージェントなどのAI技術を応用。	CPPS(Cyber Physical Production System)の構築、機能や振舞を含むモデリング技術、マルチエージェント技術、製造知識のデジタル化技術の開発が進められつつある。	装置エージェントを連携してマルチエージェントシステムとして仮想生産ラインを構成する技術の確立や、実運転データとシミュレーションデータを利用した最適な生産システムの運転制御技術などが開発される。	実績データに基づいて、製品に合わせて最適な製造を実施するCPPSベースの生産システム構成が定着する。これにより、工場同士が連携し、広く製造知識も含む情報共有を行なう仕組みもできる。
• 製造装置の自律化・知能化 工作機械、産業ロボットなど製造対象に対して自動的かつ自律的に最適な動作で対応したり、装置自身が予防保全や協調動作を自律的にこなす技術。ビッグデータ(運転実績データ)の利用、パターン認識、機械学習、生産技術知識表現などのAI技術の応用。	主にプログラムにより動作が決まる装置について、これまでの作業データを利用した機械学習や深層学習、パターン認識技術などにより、より高度な最適化、自動化、自律化実現する技術の開発が進められつつある。	装置自身で状態データを収集し、そのデータと与えられる製造対象物のデータを利用して自身の判断でより最適な装置の運転をする技術の開発などにより、さらに、個々の装置の自律化・知能化がさらに進む。	さらに知能化が進み、製造装置が自身の機能・振舞いも含んだモデルを自身の中にもち、Cyber-Physical一体装置とする技術が開発され、装置自体がCPPSの主要な構成要素となる。
• インテリジェント製品と開発支援 従来製品と情報技術の融合により、より快適で安心安全な日常生活を支える製品の出現と、その製品の開発・設計と製造準備を支援する新しいCAD/CAEやCAMシステムなどの開発。ビッグデータ(設計データ)の利用、プランニング、機械学習などのAI技術の応用。	自動運転車、介護ロボット、デジタル家電などいわゆる人工知能搭載の日常生活製品の開発が始まっている。一方、CAD/CAE/CAMなど製品の開発支援技術は従来の考え方のままで構成している。	機械学習、パターン認識など人工知能技術をソフトウェアとして一般の製品に組み込む技術の確立により、顧客の個別希望に合わせたカスタマイズ製品の実現が可能となる。顧客個別注文支援CADが開発される。	顧客が、欲しいもののイメージを入力すると、世界中の製品情報を参照して自動設計する技術や、CPSSと連携し製造できる工場へ繋ぐ技術が開発され、製品のオーダーメイド支援環境が整えられる。

ネットワーキングの発展動向

ネットワーキングとは、知識や情報を共有することにより人間社会を豊かにするための技術。

発展動向の整理軸	現状の動向	5年後の動向	10年後の動向
基盤技術 情報共有, 同期／非同期コミュニケーション, 共有仮想環境, ナレッジマネジメント, ゲートウェイインタフェース, セキュリティ	クラウドコンピューティング ユビキタスコンピューティング ソーシャルネットワーキング グループウェア	携帯端末(スマートフォン等)やクラウドを活用したコラボレーションシステム サイバーセキュリティ対策	新しい情報通信技術の出現
適用範囲 企業活動、教育システム、社会活動	企業内利用(会議支援, 協調作業支援) 教育・学習支援 電子商取引(eコマース) IoTと人工知能との融合	分散型企業活動 遠隔教育⇒教育制度の変革 オープンサイエンス／研究データ等の蓄積と保存 ネットワークサービス(医療, 福祉, 防災, 行政等)	分身が活用可に
社会の成熟度 情報／コンピュータリテラシー 人間社会の知的な活動	SNSの弊害 コピーアンドペーストの蔓延 倫理観の欠如 個人情報の漏洩と過剰保護	倫理教育(研究倫理, 技術倫理, 生命倫理, 等々)の浸透 個人情報を保護しつつ、適切な活用が可能になる	社会の知的化

社会取引・フィンテックの発展動向

フィンテックとは、“finance”と“technology”を組み合わせた造語であり、2010年以降、IT技術を応用した様々な新しいサービスが提供されつつある。対象は、「決済」、「融資」、「海外送金」、「PFM(個人向け資産管理)」、「クラウド会計」、「投資」、などの金融関係のサービスから、さらには、執行機能を持つ契約書のIT化や組織の形成を自動化するDAO(分散型自動化組織)など、社会取引全般に及ぶ。ITによって既存サービスの効率化を促進するだけでなく、リアルタイム型に顧客と一体になり新たな付加価値を生成する仕組みも内蔵できるため、金融機関のみならず、企業のあり方を大きく変貌させる潜在力がある。

発展動向の整理軸	現状の動向	5年後の動向	10年後の動向
・会計業務ソフト 個人向け銀行口座管理や自動家計簿、企業向けの経理・財務管理・税務書類作成、国をまたがる国際会計、などお金の出入りに関する様々な複雑な処理を自動的に行う技術。	個人向け資産管理、中小企業向けの会計システムがクラウドコンピューティングによって廉価に提供されるようになり、経理の効率化が進められている。	多くの企業が会計業務ソフトを使用ようになる。オープンAPIが進み、さらに高度化したサービスの提供が可能となる。商流情報も活用されるようになり、データに基づく融資先の信用リスク評価も可能となる。	個人・企業ともにクラウド会計が標準的になり、クラウドの中に蓄積されたビッグデータを活用した新しいサービスが次々と提供される。
・ブロックチェーン技術 分散型のコンピュータネットワークにデータを共有することで、廉価に高速に信頼性の高い取引を可能とする技術。サイバーセキュリティと連動して、新たな社会的信用の基盤となる。	ビットコインに代表される仮想通貨は既に市場に出回り、技術的な不安はあるものの、取り扱う店舗数も増大している。大手金融機関も、大規模に実用化した場合の検証実験を行うなど、潜在的な可能性を探っている。	現状では高額の手数料がかかる海外送金が廉価に高速に処理できるようになる。現実の通貨と同等の信用性を持つような仮想通貨が流通するようになる。	国家の財政破綻によって信頼が弱まる既存の通貨の弱点を補えるような安定的な価値を持つ国際的な仮想通貨が流通する。 海外など遠隔地とのビジネス契約をIT技術で簡単に短時間で実行できるようになる。
・サービスデザイン 出来上がったサービスを売るのではなく、顧客とともに新しいサービスを創造しながら発展する形の新しいビジネス形態。背後には、顧客のニーズを自動的に分析するビッグデータ解析とAI、リスクを評価するシミュレーションモデルなどの開発が必要である。	ロボアドバイザーが、顧客ごとのリスク嗜好に応じた資産運用案を提案できるようになり、簡単なファイナンシャルプランナーの仕事に代用できるようになっている。	現状では消費者金融などが担っているような小規模の融資の判断をAIが行うようになる。クラウドファンディングが広まり、普通の人々が少額からの資金運用を手軽に行えるようになる。	分散型自動化組織が実用化され、目的と狙いに賛同する人を集めて契約を交わして組織を稼働し、組織の運営によって得られた利益を自動的に分配するような新しいタイプの企業が誕生する。

行政ミクロ情報等の利活用の発展動向

関連法制の改正を通じて、政府・自治体が収集している統計作成のためのマイクロデータや行政情報の政策立案・モニタリング・評価への利活用を促進するためのデータ基盤が確立するとともに、そのデータ基盤を全府省・自治体・全国の研究機関でセキュアに利活用する研究・政策検討環境が整備される。また、それらを支えるデータサイエンティストの系統的育成を通じて、行政ミクロ情報と民間保有のビッグデータとのデータ融合活動を通じた公共政策の質の向上が進展する。行政ミクロ情報の民間による利活用の仕組みや国際標準化も進展し、データサイエンティストによるデータサイエンスに基づく社会改革が国内外で進展する。

発展動向の整理軸	現状の動向	5年後の動向	10年後の動向
・EBPM (Evidence Based Policy Making) EBPM,証拠に基づく政策立案は、政策の前提になる統計、統計マイクロデータ、政策への活用に資する行政情報客観的なデータに基づいて、公共政策に関するPDCAサイクルを適切に回す活動。 ・オンサイト施設 官民の秘匿性の高いミクロ情報を管理する中央施設とネットワークで繋がったシンクライアント環境上で、情報のダウンロードあるいは閲覧をすることなく、モデリングしたり探索的データ分析を可能とするデータ分析環境 ・データ融合(Data Fusion) 無作為抽出された公的統計マイクロデータと有意抽出された民間ビッグデータ、あるいはアクセスパネルデータを結合して、より母集団に対する妥当なモデリングを可能とする統計的推測技術。	米国には、EBPMの進め方、それを支援する統計のあり方を議論する政府委員会が存在する。わが国でも、EBPMを府省横断で推進する総括機能の設置が議論されている。 海外では、政策研究を定められたオンサイト施設や有研究者の研究室からの接続を許可など、様々なオンサイト施設が設置されている。わが国では、総務省が統計作成で収集したマイクロデータを利用できるオンサイト拠点の試行運用が2017年より開始されている。 Propensity Scoreに基づく因果推論などの研究は進展しつつあり、個別応用は進んでいる。統計マイクロデータとPOSデータとを融合するプロジェクトも提言されている。	各府省にEBPM統括官、府省横断的EBPM委員会が設置され、EBPM支援のための統計ないしは情報の基盤整備と行政データサイエンティスト層の系統育成拠点が整備される。 各府省・自治体がEBPMの推進と社会科学・政策科学など公益性の高い研究に対して行政のミクロ情報等に基づく探索的データ研究を可能とするリモートアクセス型拠点が全府省、全国に設置される。 母集団からの計画的抽出に基づく公的統計マイクロデータ、レセプトなど悉皆的行政情報、有意抽出された民間ビッグデータの融合が体系的に開始され、公的統計、民間情報分析の質向上に寄与する。	公共政策、地域政策におけるEBPMのグッドプラクティスが多数生成され、経済成長や国民生活の向上への寄与も可視化される 秘匿処理、秘密計算技術の進展により、データ閲覧可能なりリモートアクセス拠点だけではなく、一般の環境からも、一定のデータ・サイエンスに関する力量を有する者が、データ閲覧は不可能だか行政ミクロ情報に基づく公共性の高い計量モデリングが可能となる。 政府・民間情報の融合を加速するメタデータ国際標準などが確立し、世界義母での必要なデータ融合が加速する。

人工知能リスクマネジメントの発展動向

人工知能(AI)リスクマネジメントとは、AI技術の適用に伴って好ましくない結果が発生するリスクに対して、事前にリスクを評価して適切な対策をし、好ましくない事態の発生後に適切な対応をすること。

発展動向の整理軸	現状の動向	5年後の動向	10年後の動向
・技術的対策の高度化 機械学習したAIが誤った回答をするリスクとその社会的影響 誤った回答への技術的対策に関する研究開発の例	意思決定に利用する局面 ・AIの誤りが散見される ・人が気付いたり代わって対応可能などで、影響は限定的 人がAIを理解することを支援 ・AI自身の説明能力 ・学習範囲・結果の提供 ・シミュレーション環境	AIが自動的に動作する局面 ・AIの誤りが人に実害を与える。人の意図しない動作をする ・利用者の損害が発生するが、影響は特定AIの利用者に限定 人がAIを評価・選択することを支援 ・AI評価基準の設定 ・AI自身や第三者による評価 ・AI適用範囲の制御	AI間で自動連携する局面 ・誤りによる被害の範囲と規模が拡大する ・影響が、社会的に広がる可能性がある AIと人のロバストなネットワーク ・AIの自己品質保証 ・AI同士の信頼性確認 ・事故原因の特定、切り分け
・社会的対策の高度化 AIがもたらす好ましくない社会的影響（不安、リスク） 社会的対策の例	AIの期待が大きい局面 ・人の職が奪われる不安 ・制御不能になる不安 開発者視点 ・『AI開発原則』 総務省 ・『人工知能学会 倫理指針』 ・『官民ITS構想・ロードマップ 2016』 ・"Ethically Aligned Design", IEEE	AIの限界が認識される局面 ・能力不足のAI ・悪意を持つAIによる犯罪 提供者視点 ・AI提供ガイドライン ・AIによる事故・犯罪情報・ベストプラクティスの共有 ・AIの仕様・責任の規格制定 ・AI事故・犯罪の警察的機能（見張り、広報、取り締まり）	AIと利用者が共存する局面 ・使いこなせない人が多数発生 ・被害・苦情対応が不十分 利用者視点 ・教育、支援、監視、緊急対応など既存社会機能のAI対応 ・PL保険や賠償責任保険などのAI対応 ・セーフティネットの構築
継続的なリスクマネジメント(リスクの分析、評価、重大なリスクへの対策)			