

### 1. 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遍在化によって大規模データ取得が実現し、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情報技術が要求されるため、ビッグデータ利活用に向けた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大きく立ち遅れる恐れがある。

2014年に当分科会から発出した提言「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、

- ①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。
- ②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。
- ③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。
- ④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。
- ⑤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。
- ⑥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。
- ⑦モバイル情報や個人や集団の位置情報等を活用した災害時の効果的対応。
- ⑧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、コンピュータ基のプロ棋士への完勝、自動翻訳や自動運転の飛躍的発展、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遑がない。

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ムーアの法則にみられるように、社会の予想や期待が技術発展をリードするという側面もあるからである。

### 1. 2 展望調査結果の概要

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整理軸を複数定め、それぞれの軸に沿っての発展予測を試みた。

#### 0. データ解析技術(削除) <<<調査資料全体を俯瞰したイントロとする

1. 計算基盤
2. シミュレーション・データ同化
3. 機械学習, 深層学習

4. パターン認識(画像認識,物体認識,音声認識)
5. 自然言語処理
6. ロボット
7. マニファクチャリング
8. [行動シミュレーション・エージェント]未提出
9. [ネットワークング]未提出
10. 社会取引・フィンテック
11. ビッグデータ,データベース
12. 人工知能リスク管理

発展動向の整理軸としては, ①機能の高度化, ②対象(データ)の多様化・高度化・大規模化などが比較的共通に見られたが, その他はオープン化, 可視化の高度化, 学習の自律性・高度化, 機能の個性化など技術課題によって異なる整理軸が採用されている。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する。

## 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つの整理軸: 計算基盤の高性能化のアプローチを示す「性能向上」軸, AI・ビッグデータを活用するための計算基盤(アーキテクチャ)の高度化のアプローチ「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(もののインターネット化, Internet of Things)などセンサーと連携した計算基盤全体の構築技術「計算基盤構築技術」軸に沿って整理する。

#### 現状の動向

「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

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メモリスループット(Bytes/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, 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 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 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(Cyber-Physical System)が現実のものとなることが予想される。

#### (2) 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の意味では, 研究対象の現象を記述する支配方程式を空間・時間とともに離散化し, 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、やはり狭義同様、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。本解説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必ず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。また、そもそもシミュレーションモデル自体が、支配方程式であったとしても対象の一種の理想型であり、現実の問題を取り扱う際には、そのモデル利用の適性についても丁寧な検討が必要である。

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。この考え方自体に特段の新規性があるわけではないが、大規模数値計算がスーパーコンピュータの発達とともに“ある程度”ストレスなく実行可能になってきたことと、大量な観測・計測データの取得と保存を可能とするシステムや基盤が整備されたことにより、1990年代以降、データ同化の研究が特に気象・海洋分野を中心に盛んになった。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はより広範な分野に広まり、特に地球・宇宙科学以外の、比較的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などである。

前述したように、データ同化はシミュレーションと観測・計測技術を融合する技術である。これまでこの両者は研究開発現場でも別個に行われがちであったこともあり、統合効果には大きな期待が寄せられている。わかりやすい例は、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効果的利用である。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である。また、現在の人工物の設計においては、CAE(Computer Aided Engineering)のみで行われることも多くなってきたため、その弱点を補強する意味でもスマートデザインのコンセプトは有用である。

これらの技術の延長上には、非常に魅力的な応用領域、または有望な市場が待っている。IoT(Internet of Things)の発展・整備により、大規模な人工物、例えば社会インフラからは膨大なセンサー情報が得られる。これらの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は、どうしても得られる情報の領域や精度に限界があるが、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など、治療を高度化させることも可能であろう。これらの技術は10年後には達成されていると思っている。

### (3) 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、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・手法のことであり、具体的には、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關係やパターンを抽出し、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。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。

これら技術の向こう10年間の発展動向を見通す上で、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。

#### ① 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

#### ② 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

#### ③ 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多様化

上記の①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。今後数年でシミュレーションを用いて仮学習し、実少数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、かつ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。10年程度先には、時空間的構造を持た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、より広範な問題適用が進むと考えられる。

②については、現状、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・複雑な学習を自律に行えるようになりつつある。この方向の発展は今後も続き、広範な実問題への適用が進むと共に、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。

③については、現状、予兆発見・検知や因果推論、クラウド技術、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が、数年後にはより多くの他の情報処理問題と機械学習原理の融合・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年程度を見越せば、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ていくと考えられる。

### (4) パターン認識(画像認識,物体認識,音声認識)

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。そのゆえ、この「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。

#### 現状の動向

「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。「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性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。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。「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、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」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る。「機能の個性化」については、ユ

一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。

#### (5) 自然言語処理の発展動向

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

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と大規模なタグなしコーパスの利用により，かなりの解析精度を達成することが可能になっている。リソースリッチな言語からリソースプアな言語へタグ付き情報をトランスファーすることによる手法などが研究されている。述語項構造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。

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。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の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。

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

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。

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。

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

文脈を考慮した自然言語解析技術が進展する。文脈の意味を表現する技術の研究が進むと考えられる。文脈を考慮した文の意味解析技術の研究が進み，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。

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。

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。

#### (6) ロボットの発展動向

AI・ビッグデータ利活用の中でも、自動認識に関するところでは、機械学習、特に深層学習手法が(最終結果の)検出や識別性能で革新的な進歩を導き、様々な状況で実際に役にたつレベルに近づいて来ている。途中の過程がブラックボックスでわからなくとも、この結果が(さえ)良ければ有効に働く分野の代表として「ロボット」が挙げられる。

例えば、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い。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。この考え方に基づき、RGB-D(カラー(RGB)と距離(Depth))画像上で把持を行う場所を矩形ウィンドウで表し、把持が可／不可のラベル付けした学習データを用いて、RGB-Dパターンと把持可能性を直接結び付けて学習する手法が提案されている<sup>1)</sup>。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。これにより、対象物ごとのモデル化が必要なく、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た。

また、Amazon Picking Challengeの課題のように、もう少し高度に、対象物を識別して取ってくる課題においても、やはり、深層学習を用い、入力のRGB-D画像から対象物認識と吸着ハンドのための吸着位置検出までをニューラルネットワークで学習する手法が成果を挙げている<sup>2)</sup>。透明ボトルや光沢物、丸いものから平たいノートまで、これまで単一の「認識→動作計画算出」手法では取り扱えないと思われていたものが、比較的単純な動作タスク「(識別して)取り上げる」であれば、深層学習とハンドの工夫で実現できる可能性が示されている。ただし、実際に行われているアクションは2指ハンドでつまむ、面を吸着して持ち上げるなど、まだ単純なものに限定される。例えば、ねじ回しを持ち上げてねじを締める、はさみを取り上げてはさみを使う、など道具の使用となるとまた、動作を行うために認識すべき事項が急激に増え、学習とその他の技術の組み合わせやさらに複雑な学習が必要となる。

機械学習の活用をさらに加速する動きとして、学習方法自体を高度化する研究も進められている。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。これを回避する方法として、シミュレーション結果を用いることも考えられているが、複雑な過程であればシミュレーションで精度良い学習データを得るのは難しい。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<sup>3)</sup>。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、人を凌駕できることを実際に示したものであり、インパクトの大きな成果である。



経産省も AI ロボット・人工知能を組み込んだロボットを開発に意欲を示し、産学官で実用化を目指し、研究拠点の整備を開始している。AI・ビッグデータ利活用を最も早く産業展開するのはロボット分野かもしれない。

参考文献 <<< 後ろに移動の予定

- 1) I. Lenz, H. Lee, A. Saxena: “Deep Learning for Detecting Robotic Grasps”, The International Journal of Robotics Research 34 (4-5), 705-724, 215.
- 2) 藤吉弘亘, 松元叡一, 岡田慧, “[特別講演] Amazon Picking Challenge 2016 の参加レポート”, パターン認識・メディア理解研究会, vol. 116, no. 461, pp. 123-129, 2017.
- 3) Google Research Blog:  
<https://research.googleblog.com/2016/03/deep-learning-for-robots-learning-from.html>

#### (7) マニュファクチャリングの発展動向

マニュ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。3つの観点から今後の動向について述べる。

【バーチャルマニュファクチャリング】 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である。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。5年後には、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の開発が進むと予想される。10年後には、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 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あがる。

【製造装置の自律化・知能化】 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に行なう技術である。現状では、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。5年後には、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。10年後には、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】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のことである。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年後には、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。顧客個別注文支援 CAD が開発され、10年後には、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。

#### (8) 行政マイクロ情報等の利活用の発展動向

2017年5月の官房長官をヘッドとする公的統計改革推進会議の報告とりまとめが、2017年6月の日本政府骨太方針に反映することで、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と制度との整備が2018年以降に実施されることが予想される。また、新しい家計インデックス(CTI)の2018年提供を目指し、総務省統計局では民間ビッグデータと統計マイクロデータ情報との融合を支える産官学連携組織の形成を開始した。これは今後加速するであろう先端統計的方法を用いた、標本調査論に基づく統計マイクロ情報と民間ビッグデータとのデータ融合の嚆矢となり、今後の官民への波及が期待される。

更に、各府省には2018年以降、証拠に基づく政策立案(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こととなった。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っている。この中で、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求められており、そのための法整備も2018年には開始される予定となった。

オンサイト施設とは、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行政マイクロ情報を分析可能とする拠点だが、そこで行ったデータ分析結果の外部持ち出しに関する審査を担当する予定の各府省、(独)統計センターにおける審査指針、審査体制を検討するために、2017年1月から総務省統計マイクロデータを提供した試行運用が一橋大、神戸大、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。2018年には、和歌山に設置される統計利活用セ

ンターを含む全 10 オンサイト施設による本格運用が予定されているとともに、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。

今後 5-10 年以内に、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、10 年を目途に、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近い、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。

#### (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。

##### ① 発展動向の整理軸

次の 3 つを、ネットワーキングの発展動向の軸として、整理する。

##### (1-1) 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、セキュリティなどの基盤技術の進展。

##### (1-2) 適用範囲

企業活動、教育システム、社会活動など、適用範囲の拡大。

##### (1-3) 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。

##### ② 現状の動向

##### (2-1) 基盤技術

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。

##### (2-2) 適用範囲

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、導入が進んでいる。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。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。単純作業の機械化は、かなり進んでいるが、まだ、第 4 次産業革命は始まったばかりである。

### (2-3) 社会の成熟度

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。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。コピーアンドペーストの蔓延も社会問題となっている。人間社会において倫理観が欠如してしまっているようである。個人情報の取り扱いについては、過渡期である。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ことも、問題である。

### ③ 5 年度の動向

#### (3-1) 基盤技術

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。

#### (3-2) 適用範囲

分散型企業活動は、技術の進歩に応じて、さらに進んでいる。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、内容の良い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を作成されるようになる。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。授業にはそれを活用し、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。

#### (3-3) 社会の成熟度

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)が浸透する。個人情報についても、社会的ノウハウの蓄積が進み、適切な活用が可能になる。たとえば、就職活動も、求職者と求人者のマッチングの高機能化が進む。

### ④ 10 年後の動向

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。一般社会において、ネットワーク技術と、IoT(センサ情報)や人工知能(ボットやエージェント)との融合が進み、第4次産業革命が急速に浸透し始めている。それにより、社会の知的化が進む。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) 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。

この分野では極めて多様な展開が今後 10 年間に想定されるが、以下の 3 点の軸でまとめることができる。

#### ① 会計業務ソフト

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

#### ② 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

#### ③ 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

上記の①については、現状、個人向けの資産管理や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っている。近い将来、様々な形で蓄積された商流情報が活用され、例えば、データに基づく融資先信用リスク評価も可能になると期待され、より高度化したサービスの提供が可能となる。10年程度先には、世界的にクラウド会計が標準的になり、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と考えられる。

②については、現状、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。大手金融機関も潜在的な可能性を探っており、近い将来、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と期待される。さらに、お金のような数値の交換だけでなく、10年程度先には、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、海外など遠隔地と IT 技術で簡単に短時間に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可能性がある。

③ については、現状、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を代用できるようになっている。近い将来、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可能性が高い。10年程度先には、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可能性がある。

### (11) ビッグデータ、データベースの発展動向

インターネットを中心とする通信技術の発達、観測・測定機器の進歩により、超巨大データが蓄積されている。また、最近の国内外の情勢から大規模問題の解決が必要とされる分野には多くの新分野が加わり、その範囲も地球的規模に拡大しつつある。しかしながら、近年のデータサイズの急速な増大はハードウェアやアルゴリズムの進化の速度を大幅に上回っている。このようなビッグデータを対象とした大規模問題群の解決には、従来のデータベース理論やアルゴリズム理論では対応不可能であり、新たな技術革新、アルゴリズム革新による解決が強く望まれている。たと

例えば、アルゴリズムの分野において、ランダムアクセスマシン(RAM)と多項式時間アルゴリズムというパラダイムが構築され、その下で多大な結果が出されてきた。これまでは多項式時間アルゴリズムならば「速い」アルゴリズムであると考えられてきたが、ペタスケールやそれ以上のビッグデータに対して  $O(n^2)$  時間アルゴリズムを直接適用するだけでは、計算資源や実行時間などの点で大きな困難に直面する。少なくとも線形時間、場合によっては劣線形時間や定数時間アルゴリズムが求められている。また、物理的に離れた場所に点在する巨大データを用いた大規模分散データベース上での高速処理も必須の技術である。さらに、様々なセンサーデータ、通信機器から巨大ストリームデータが日々、時々刻々蓄積されており、その蓄積データを用いて正確な意思決定を瞬時におこなう技術開発の必要性が、今後一層増大することが予想される。そのためには、リアルタイムでデータ圧縮をおこないながら、高速な検索も可能とする技術が求められている。このように、ビッグデータ時代を迎えて、ハードウェア的側面だけでなく、ソフトウェア的側面の発展が重要である。その基礎となるアルゴリズム理論の一層の整備が急務であるといえる。

このような背景のもと、本節では、1. アルゴリズムの高速化、2. ビッグデータに対するデータベース技術の革新、3. リアルタイム処理技術の高度化に関する、現状、5年後の動向、10年後の動向について、展望を述べる。より、具体的には、

1. アルゴリズムの高速化については、対象データの全体を見ることなしに、巧みにサンプリングをおこないながら、対象データに関する特性量をデータサイズに依存せず高速に、近似的に推定する技術の開発が必須である。例えば、数十億点からなるウェブグラフに対する探索や特性量計算を高速に行う技術が開発され、広く実用化されることが期待される。
2. 現在でも、ビッグデータを大規模分散環境を用いて高速にデータ解析可能になってきたが、センサーストリームデータと合わせた、リアルタイムビッグデータ処理が可能となる時代がやってくると予想される。例えば、株価取引データなどの金融データを用いたリアルタイム機械学習による投資意思決定の実用化が予想される。
3. ビッグデータリアルタイム高速処理技術を支える基盤技術としてオンライン圧縮技術が開発され、ビッグデータの高速伝送の基盤技術として定着し、広く社会に普及することが予想される。

## (12) 人工知能リスク管理の発展動向

人工知能(AI)への不安が話題になっており、研究開発と並行して、AIのリスクを考慮し対策する必要がある。本報告書ではいろいろな分野でAIの具体例や社会的インパクトが語られているが、ここでは分野によらない全般的なAIのリスクへの対策を展望する。大きく、機械学習が誤るリスクへの技術的な対策と、不安への社会的な対策の二つに分ける。

### (1) 学習誤りによるリスクへの技術的対策

ディープラーニングによるAIは、従来のITシステムのような仕様を明確に規定し、能力を保証することが困難となる。AIは、誤る可能性があること前提として利用される。利用する局面を次の三段階と想定して、技術的な対策の例を挙げる。

①AIの結果は、利用者の意思決定支援に利用される。人の判断が入る。

②AIの結果が、自動的に機械の作動や、注文の入力操作につながる。

③AIの結果が、自動的に他のAIの入力となり、AI同士が連携する。

①には、人が結果を判断できるよう、人がAIの能力を理解するための技術が望まれる。②では、AIを利用するか否かを判断できるよう、人がAIを評価・選択するための技術が望まれる。③では、あるAIが誤っても、AIのネットワークの中では発見、部分停止、修復できるようなロバストなネットワークを実現する技術が望まれる。しかし、技術的な対策だけでは不十分であり、(2)で示すような社会的な対策もあわせて必要となるであろう。

## (2) 不安やリスクへの社会的対策

“人工知能”という用語が、人に従来のITシステムにない過大な能力を想像させ不安をもたらす。AIが健全に成長するよう、不安をリスクとして捉えて評価・対策しながら利用する必要がある。AIへの不安が生じる局面を次の三段階と想定して、社会的な対策の例を挙げる。

①AIが登場し、期待が膨らむ反面、人の職がなくなるなどの不安が生じる。

②AIを利用すると、理解不能、故意・悪意が隠されるなどAIの限度、弱点の不安が生じる。

③人とAIが共存することが普通になると、AIへの不適応感を抱く人が多くなる。

①には、開発者視点で、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。②には、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、ベストプラクティスと事故・犯罪情報の共有、事故・犯罪から人を守る警察的な機能が期待される。③には、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が期待される。

AI技術の発展が人間にとって心地よい社会につながるためには、継続的に新たなリスクの抽出・評価、対策を行なうリスクマネジメントに留意する必要がある。

(13) 文献 (文献はここにまとめる)

[1] 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 提言、ビッグデータ時代に対応する人材の育成、平成 26 年 9 月 11 日。

(14) 用語集(一般読者を想定し、使用した用語のうち必要なものをここにまとめて掲載する)

(15) 付録

各技術課題の PPT を付録としてつける(?)