

理化学研究所における 人工知能研究開発の取組

2016年5月24日

理化学研究所 革新知能統合研究センター
センター長(内定者) 杉山 将

理研AIPプロジェクト について



理化学研究所

理化学研究所 革新知能統合研究センター
センター長（内定者） 杉山 将

革新知能統合研究センターについて

- 理化学研究所は、文部科学省が進めるAIP（人工知能/ビッグデータ/IoT/サイバーセキュリティ統合プロジェクト）の研究開発拠点として、4月14日付で**革新知能統合研究センター**を設置

－センター長（内定）：

杉山 将（すぎやま まさし）

東京大学大学院教授

※ 正式に着任するまでの間、
理事 松本洋一郎がセンター長を兼務

－特別顧問（内定）：

金出 武雄（かなで たけお）

米カーネギーメロン大学教授



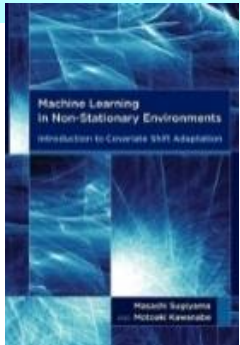
これまでの研究： 機械学習の理論と応用

教師付き学習：人間が教師となり、
コンピュータを学習させる

強化学習：エージェントが試行
錯誤を通じて学習する

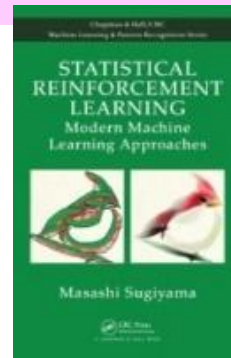
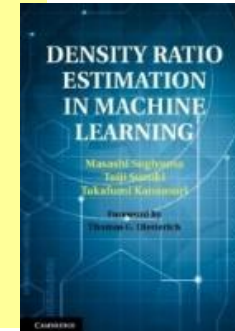
脳波によるコンピュータの操作
(独Fraunhofer研究所との共同研究)

ヒューマノイドロボットの運動制御
(NICT・ATRとの共同研究)



教師なし学習：コンピュータが
人間の手を介さずに学習する

- データの可視化
- クラスタリング
- 密度比推定



人工知能研究の位置づけ

- 工学や科学の難問を抽象化し，
数学とコンピュータを駆使して問題解決！

工学応用

テキスト，音声，画像，映像，
ロボット，広告，電子商取引

科学応用

医療，生命，物理，化学，
材料，宇宙，地学，社会

人工知能

数学

ベクトル，行列，微分，積分，
確率，統計，論理，最適化，幾何

コンピュータ

プログラミング，ハードウェア，
ネットワーク

人工知能研究の現状

- **欧米**: 巨大民間企業等が数百億～数兆円規模の莫大な予算を投じて研究開発を開始
 - Google, Microsoft, Facebook, Amazon, Toyota, OpenAI,...
- **国内**: 政府が中心となって, 数十億～数百億円規模の予算を幅広い分野に配分

国際会議NIPS2015での アジア勢の動向

- 体感的には、NIPSでの**日本人の存在感は皆無**：
 - 韓国は日本よりやや存在感があり、
中国はかなり溶け込んでいるという印象
- しかし、現在までの日本の情報科学の幅広い分野での研究の蓄積、研究者の質の高さを鑑みれば、今後飛躍をするチャンスは多いにある！

研究の動向

学習法

⋮

密度微分

異常検知

強化学習

密度比

非定常適応

分類

回帰

データ量

ビッグ

スモール

線形

加法

カーネル

深層

...

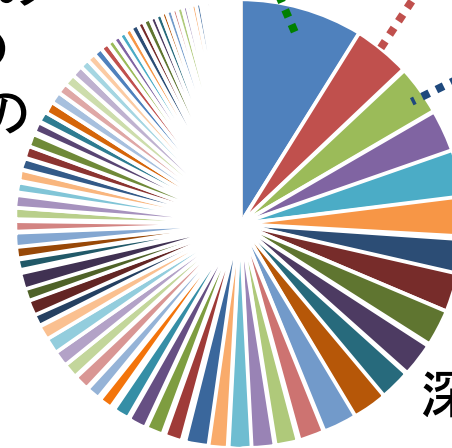
モデル

Deep Learning (9%)

Convex
Optimization
(4%)

Clustering (3%)

NIPS2015の
1838件の
投稿論文の
カテゴリ



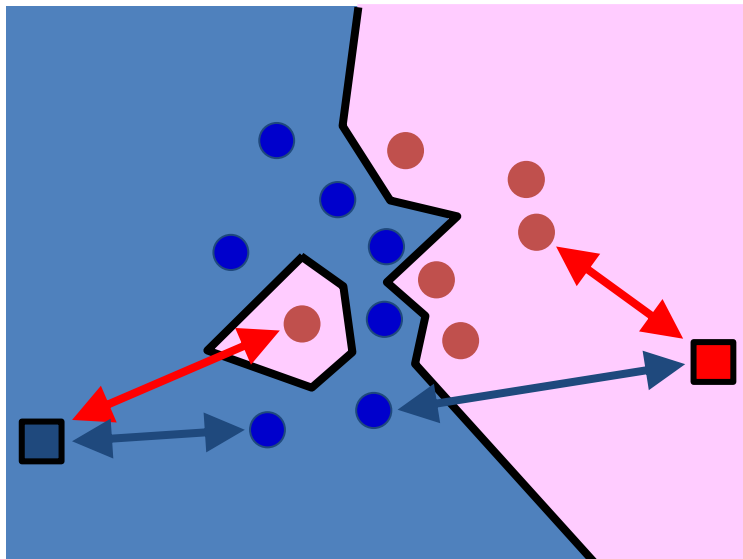
深層学習が多いが、
支配的ではない

現在の
トレンド

ビッグデータ + 深層学習 = 究極の人工知能？

- データがたくさんあれば, 古典的な
最近傍分類でもそこそこうまくいく

データ	深層学習 (CNN)	最近傍分類
印刷数字60万文字	0.01%	0%
手書き数字80万文字	0.11%	0.97%

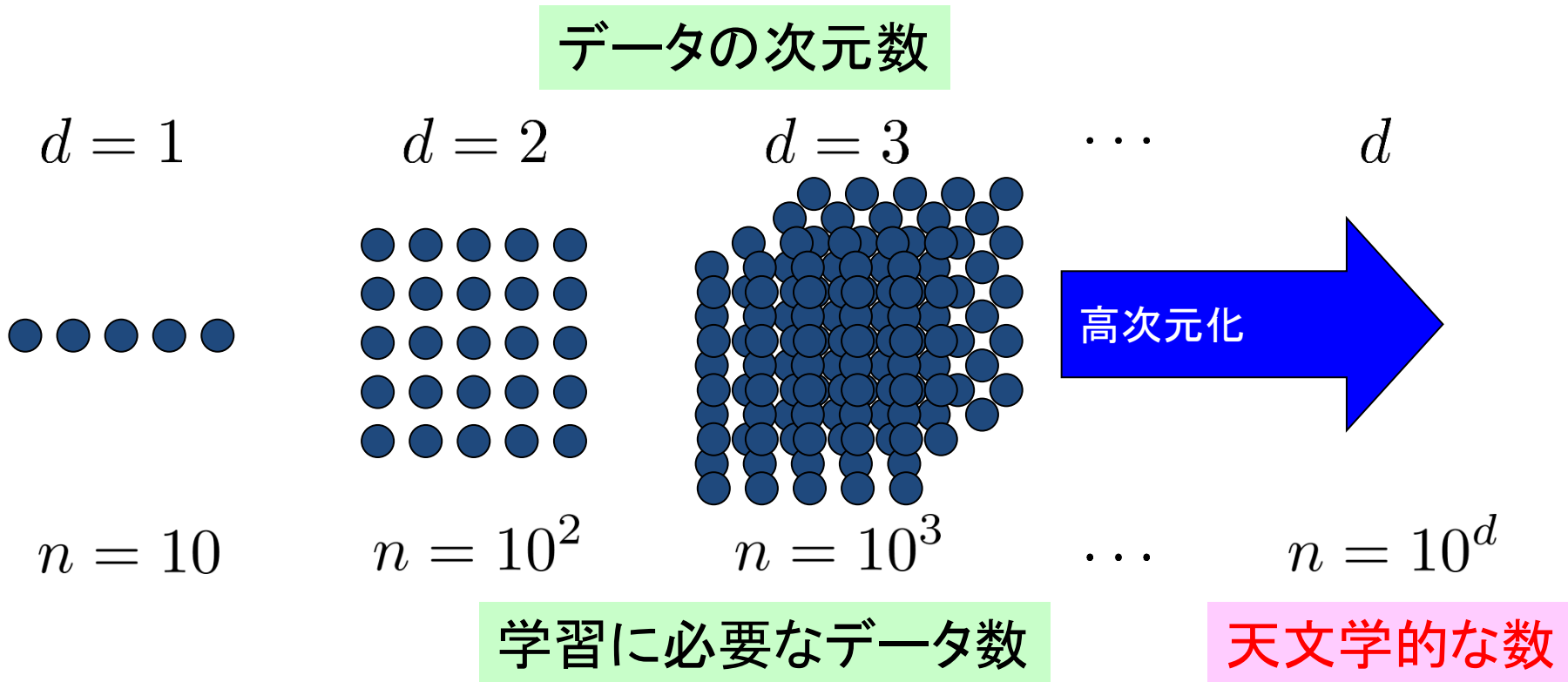


[PRMU2016/2内田ほか]

最近傍分類:
一番近くのデータと
同じクラスに分類

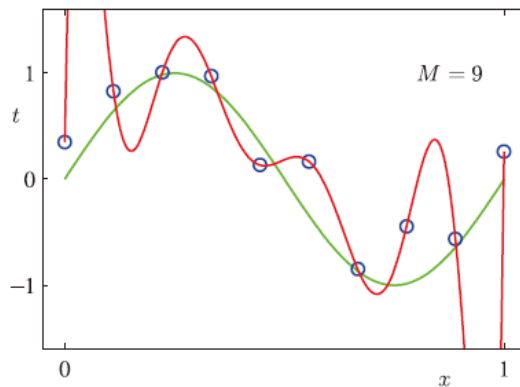
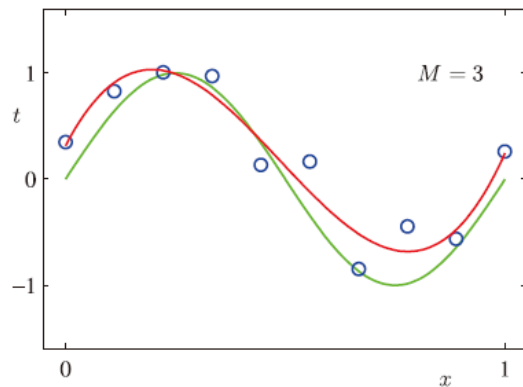
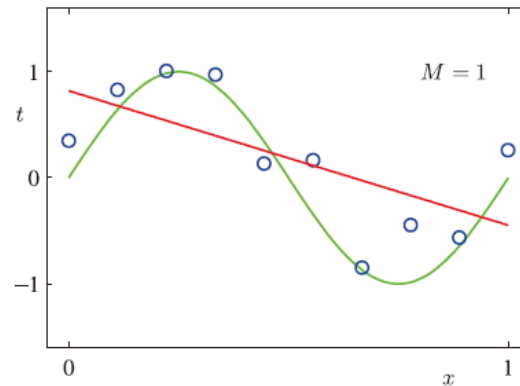
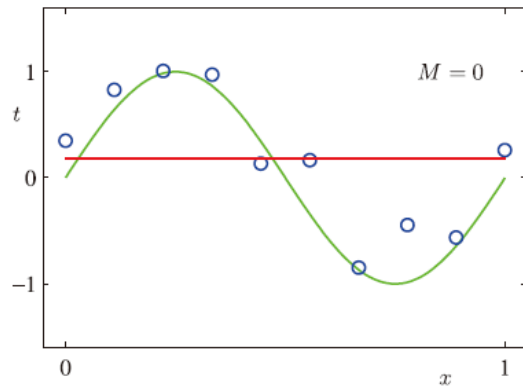
次元呪い

- センサーの数を増やすとデータの次元数が増加
 - 学習に必要なデータ数 n は, データの次元数 d に対して**指数関数的に増加**
- いくらビッグデータを集めても, **データはまばら**



究極の人工知能は汎化能力！

- **汎化能力**：教わっていないことを、過去の事例から類推する能力
- まばらなデータの間を補完・予測



[パターン認識と機械学習 Bishop 2006]

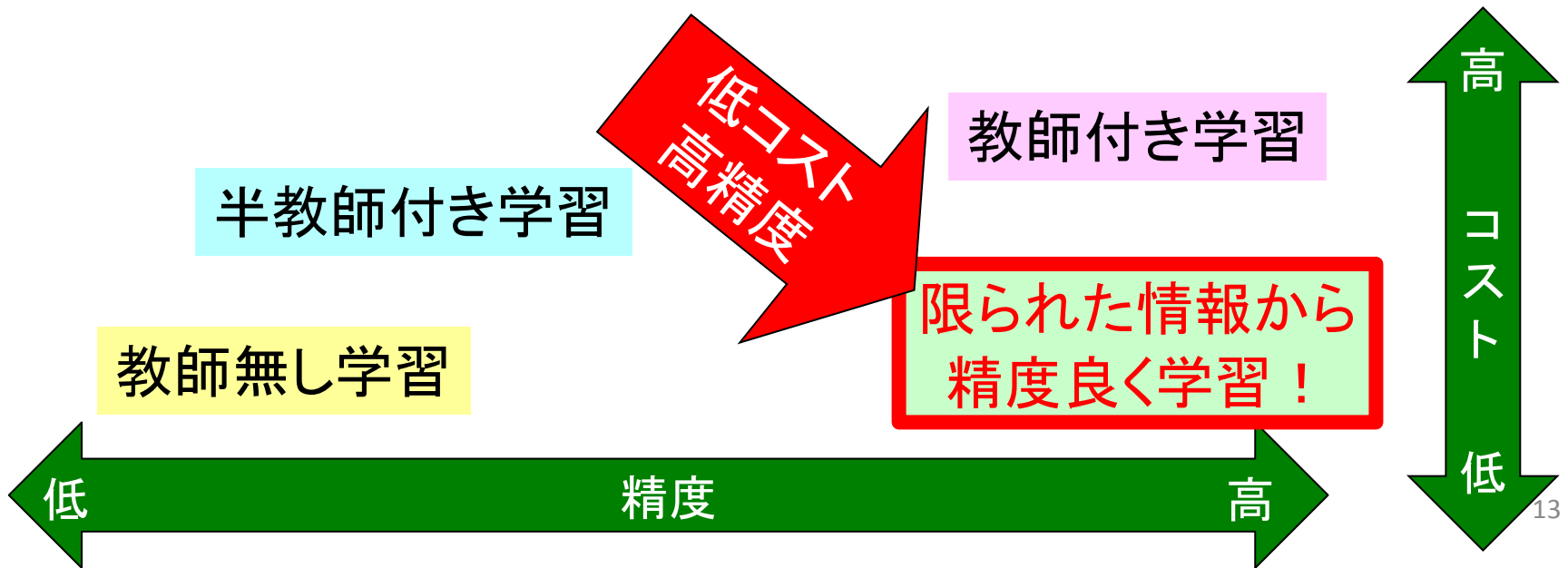
汎化の原理



- 6面体のさいころを何回も振れば、その平均値は大体3.5
- 真の期待値(答え)がわからなくても、**たくさんのデータの平均**を取れば大体答えがわかる
- **中心極限定理**: **同じさいころ**を n 回振ると、その平均値は本当の期待値に $1/\sqrt{n}$ の速さで近づく
- ビッグデータを用いれば答えが簡単に予測できそう？

ビッグデータの現実

- 同じさいころを多数回振れない
 - 例：同じ病状の患者はそれほどたくさんいない
- そもそも答えのないデータばかり
 - 例：ドローンで橋梁の画像をたくさん撮っても、危険個所は人間が特定しないといけない
- 限られた情報からの学習が重要！



AIP: Advanced Integrated Intelligence Platform Project

人工知能/ビッグデータ/IoT/サイバーセキュリティ統合プロジェクト

平成28年度予算額 : 5,448百万円 (新規)
(関連する既存事業 (2,849百万円) を含む)
※運営費交付金中の推計額含む

【国際的な動向】

- 各分野での**ビッグデータの集積、センサーの量的・質的拡大** (IoT: Internet of Things)
- **人工知能**に50年来の大きな技術的ブレークスルー (自ら特徴を捉え進化する人工知能が視野)
- 一方、**高度化する脅威に対するサイバーセキュリティの確保** (ますます巧妙化しており、人材育成が必須)

【文部科学省の対応】



- (1) 文部科学省が持つ**ビッグデータの解析** (コホート、環境のデータなど多様) を通じて、新たな価値を創造。
 - (2) そのため、**革新的な人工知能技術を開発・活用**
 - (3) ビッグデータの充実のため、高度な**センサー/IoT技術**を活用。あわせて、堅牢な**セキュリティ**を構築。
- 〔**経済産業省・総務省との連携**を呼びかけ、基礎研究から社会応用まで、一体的に実施する体制を構築〕

AIPセンター (理化学研究所) 1,450百万円

- I. 人間の知的活動の原理に学んだ革新的な人工知能の基盤技術を開発。
- II. 人工知能とビッグデータにより複数分野においてサイエンスを飛躍的に発達させる。
- III. 具体的な社会・経済価値を創造する多数の応用領域の社会実装に貢献。
- IV. 人工知能等が浸透する社会での倫理的・社会的課題等に対応。
- V. データサイエンティスト、サイバーセキュリティ人材等を育成。

戦略的創造研究推進事業 (一部) (科学技術振興機構)

新規採択課題分 1,150百万円
関連する既存採択課題分 2,849百万円
※運営費交付金中の推計額

- 大学等の研究者から広く提案を募り、組織・分野の枠を超えた時限的な研究体制を構築して、**戦略的な基礎研究を推進**。

一体的に
実施

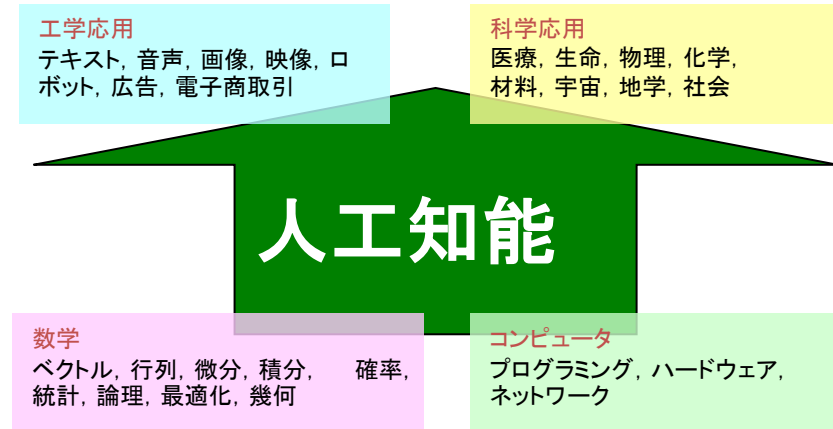
【目指す成果】

人々と社会のための知能とイノベーションの創出に向けて、世界的に優れた競争力を持つ研究者を結集、最先端研究を統合。我が国が直面する労働力減少、高齢化社会の中でも、

- ・生産性の大幅な向上による**経済成長への貢献**、
- ・一人ひとりに優しい**社会構築** (医療・介護等) を実現

理研AIP事業で行う 研究開発の考え方

- 数理科学等に立脚した人工知能の原理・原則の解明を進め、世界をリードする**革新的人工知能**
基盤技術を構築
- サイエンスや実社会などの**幅広い出口に向けた**
応用研究を進める
- 未来の科学研究に必要となる**データ構造，データ取得技術の世界標準化**を図る



理研AIP事業で行う研究開発の例

- 超高齢社会へ向けた医療サポート
 - 動画認識・センサ情報解析・アクチュエータ制御技術を融合し、高度な手術を支援
 - 機械学習による電子医療記録の自動解析・予後予測
 - 会話を通じて高齢者の認知機能を維持向上
- 老朽化が進むインフラへの対応
 - 自動操縦ドローンを駆使し、動画や打診音から橋梁などの危険個所を自律的に特定
- 甚大な自然災害への対応
 - ビッグデータ解析技術とシミュレーション技術を統合し、甚大災害をもたらす異常気象を適確に予測し、被害を最小限に抑え迅速に復旧できる社会システムを構築

理研AIP事業で行う研究開発の例

- プライバシー

- プライバシー, 説明責任, 公平性, 差別配慮, セキュリティなどを考慮した人工知能技術の開発

- 細分化が進む科学研究への対応

- 論文・特許・実験結果をもとに科学研究の発展を支援
- これまで埋もれていた発見を見逃さない技術を開発
- 次に実験すべき項目を過去のデータに基づいて決定
- マテリアルズ・インフォマティクスなどへの応用

- 倫理

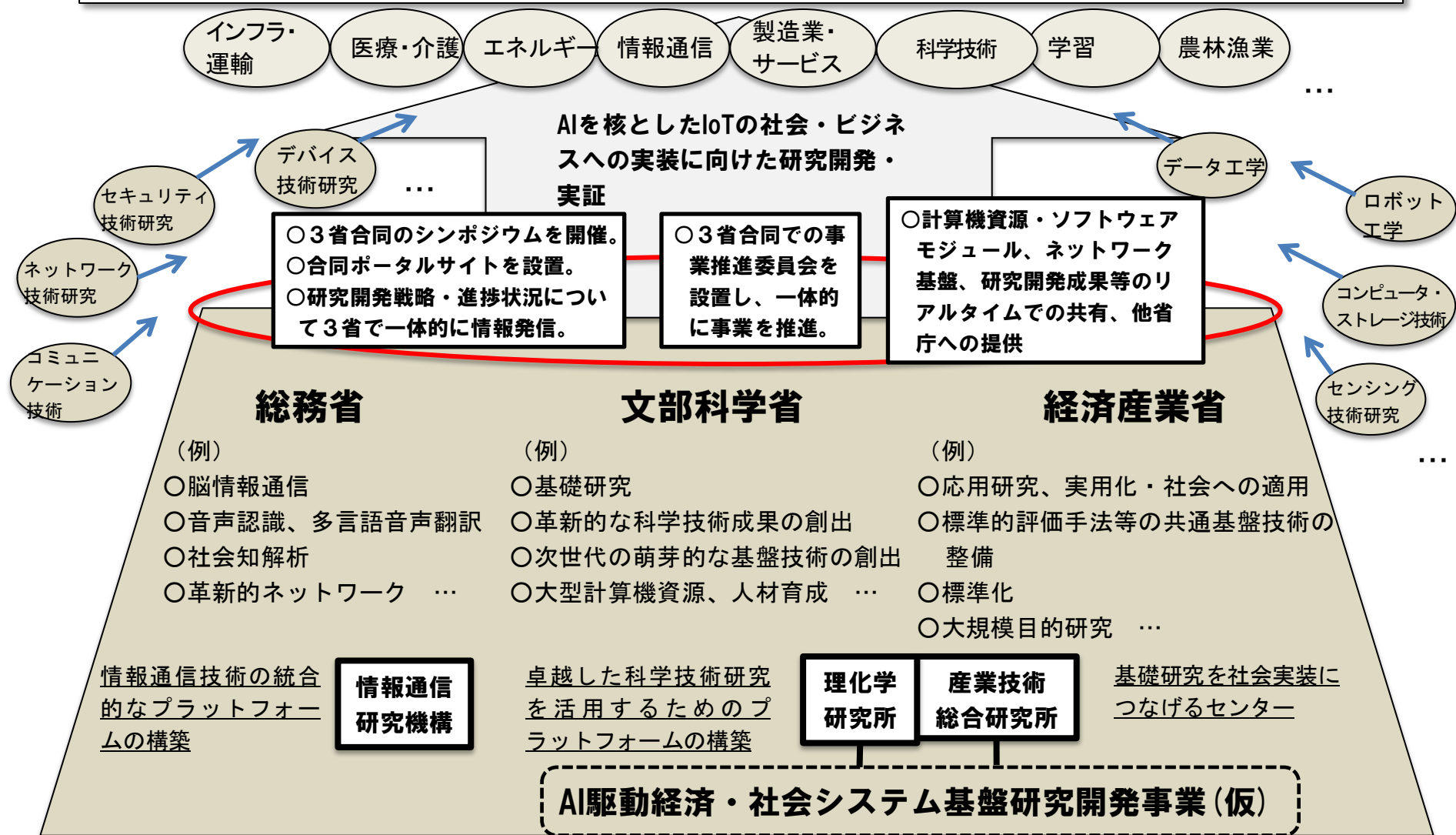
- 人工知能技術が社会に浸透する際のELSI問題や社会的影響を, 人文科学・社会科学の視点で検討

理研AIP事業で行う研究開発の例

- データサイエンティスト, サイバーセキュリティ人材等の育成
 - クラウドソーシングを用いた遠隔データ解析教育
- 革新知能プラットフォーム構築
 - 深層学習等のビッグデータを用いた学習の革新
 - 疎・不完全・超高次元データからの高精度学習の実現
 - ストリーミングデータに対するリアルタイム学習の実現
 - 学習アルゴリズムの選択・調整を自動化
 - CPU, GPU, ディスクI/O等を考慮した学習技術の実現
 - 最適なデータ収集戦略の策定

次世代の人工知能技術の研究開発における3省連携体制¹⁹

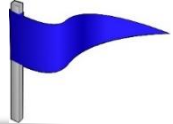
- (1) 各分野でのビッグデータの集積、センサーの量的・質的拡大 (IoT: Internet of Things)。
- (2) 人工知能の50年来の大きな技術的ブレークスルー (自ら特徴を捉え進化する人工知能を視野)。
- (3) 3省連携による研究開発成果を関係省庁にも提供し、政府全体として更なる新産業・イノベーション創出や国際競争力強化を牽引。



All Japanを超えた 国際研究拠点を目指して

- 欧米の大学・企業は「外国人」が集う国際拠点
 - 助っ人外国人を雇うのではなく、国際拠点を目指す
 - 外国人率30%以上を目指す
- Money: That's What I Want ?
 - 国際的に競争力のある待遇で有力な研究員を雇用
- 欧米と比べて、日本の大学では博士課程に進学する学生が非常に少ない
 - 修士・博士課程の学生のサポートを充実
 - 研究員に長期のポストを提供
 - 企業・大学と連携し、人材の流動性を高める

日本は欧米の周回遅れ? → No!

- 欧米は欧米のやり方で山頂を目指している
- その道を日本が今から追いかけても、 追い越すことは難しい
- 日本は日本のやり方で山頂を目指すべき!



理研AIPプロジェクト について

2016年4月14日

理化学研究所 革新知能統合研究センター
センター長(内定者) 杉山 将

革新知能統合研究センターについて

理化学研究所は、文部科学省が進めるAIP（人工知能/ビッグデータ/IoT/サイバーセキュリティ統合プロジェクト）の研究開発拠点として、4月14日付で「革新知能統合研究センター」を設置した。

1. 名称：

革新知能統合研究センター

2. センター長（内定）：

杉山 将（すぎやま まさし） 東京大学大学院教授

※ 正式に着任するまでの間、理事 松本洋一郎がセンター長を兼務

3. 特別顧問（内定）：

金出 武雄（かなで たけお）氏 米国カーネギーメロン大学教授

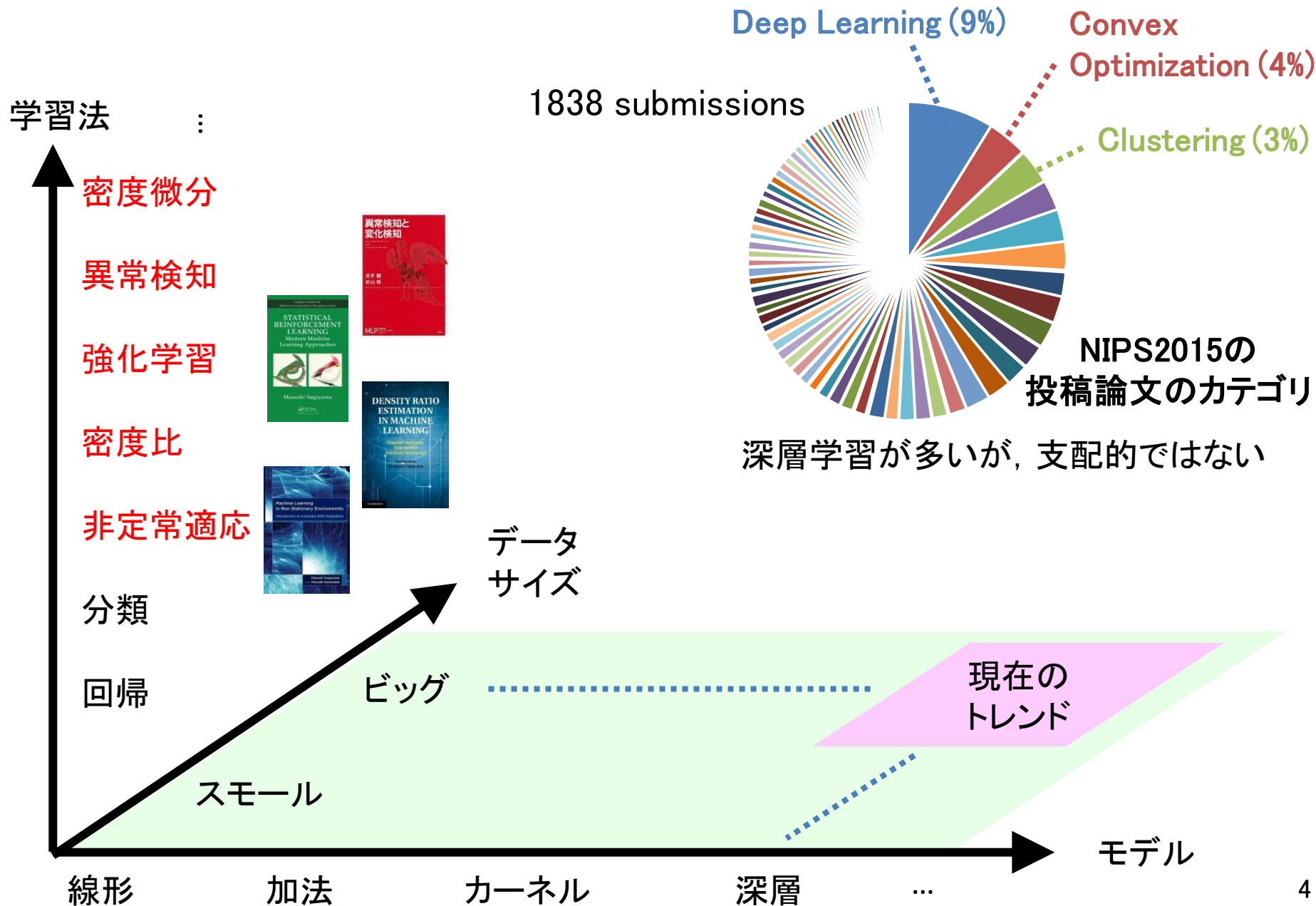
現状の人工知能研究

- 欧米: 巨大民間企業等が数百億～数千億円規模の莫大な予算を投じて研究開発を開始。
 - Google, Microsoft, Facebook, Amazon, Toyota, OpenAI,...
- 国内: 政府が中心となって、数十億円規模の予算を幅広い分野に配分。
- 欧米の研究の後追いをしても勝ち目がない。

国際会議NIPSでのアジア勢の動向

- 中国、韓国は欧米の大学・企業に所属している学生・研究者が多数いる。
- 体感的には、国際会議NIPSでの日本人の存在感はほぼゼロ、韓国は日本より少しマシ、中国はかなり溶け込んでいるという印象。
- しかし、現在までの日本の情報科学の幅広い分野での研究の蓄積、研究者の質の高さを鑑みれば、今後飛躍をするチャンスは多いにある！

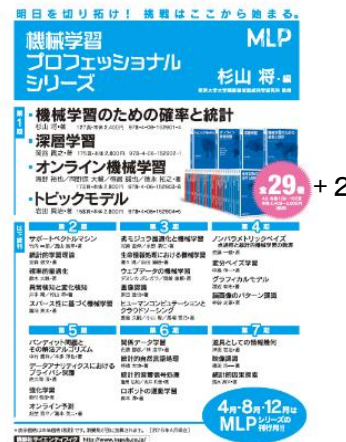
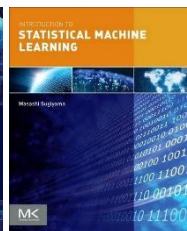
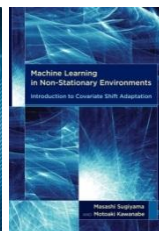
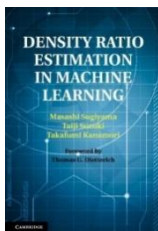
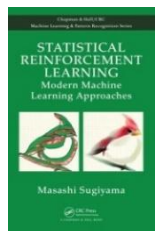
研究の動向



研究分野と活動



- 知能情報学分野における機械学習の理論と応用
- 英語論文約150編, 国際会議論文約130編, 著書9冊(英語4冊):
 - Neural Computation (NeCo) 21編
 - Journal of Machine Learning Research (JMLR) 12編
 - Machine Learning (MLJ) 9編
 - Neural Information Processing Systems (NIPS) 19編
 - International Conference on Machine Learning (ICML) 16編
 - Artificial Intelligence and Statistics (AISTATS) 10編
- 国内外の大学・研究所・企業にて約170件の招待講演
- 学会や財団法人等から28件の表彰
- 機械学習分野最大の国際会議Neural Information Processing Systems (NIPS)にてアジア人初のプログラム委員長・実行委員長
- Asian Conference on Machine Learningの共同創設者
- 31冊にわたる機械学習の教科書シリーズの監修



AIP: Advanced Integrated Intelligence Platform Project

人工知能/ビッグデータ/IoT/サイバーセキュリティ統合プロジェクト

平成28年度予算額 : 5,448百万円 (新規)
(関連する既存事業 (2,849百万円) を含む)
※運営費交付金中の推計額含む

【国際的な動向】

- 各分野での**ビッグデータの集積、センサーの量的・質的拡大** (IoT: Internet of Things)
- **人工知能**に50年来の大きな技術的ブレークスルー (自ら特徴を捉え進化する人工知能が視野)
- 一方、**高度化する脅威に対するサイバーセキュリティの確保** (ますます巧妙化しており、人材育成が必須)

【文部科学省の対応】



- (1) 文部科学省が持つ**ビッグデータの解析** (コホート、環境のデータなど多様)を通じて、新たな価値を創造。
- (2) そのため、**革新的な人工知能技術を開発・活用**
- (3) ビッグデータの充実のため、高度な**センサー/IoT技術**を活用。あわせて、堅牢な**セキュリティ**を構築。
〔**経済産業省・総務省との連携**を呼びかけ、基礎研究から社会応用まで、一体的に実施する体制を構築〕

AIPセンター (理化学研究所) 1,450百万円

- I. 人間の知的活動の原理に学んだ革新的な人工知能の基盤技術を開発。
- II. 人工知能とビッグデータにより複数分野においてサイエンスを飛躍的に発達させる。
- III. 具体的な社会・経済価値を創造する多数の応用領域の社会実装に貢献。
- IV. 人工知能等が浸透する社会での倫理的・社会的課題等に対応。
- V. データサイエンティスト、サイバーセキュリティ人材等を育成。

戦略的創造研究推進事業 (一部) (科学技術振興機構)

新規採択課題分 1,150百万円
関連する既存採択課題分 2,849百万円
※運営費交付金中の推計額

- 大学等の研究者から広く提案を募り、組織・分野の枠を超えた時限的な研究体制を構築して、**戦略的な基礎研究を推進**。

一体的に
実施

【目指す成果】

- 人々と社会のための知能とイノベーションの創出に向けて、世界的に優れた競争力を持つ研究者を結集、最先端研究を統合。我が国が直面する労働力減少、高齢化社会の中でも、
- ・生産性の大幅な向上による**経済成長への貢献**、
 - ・一人ひとりに優しい**社会構築** (医療・介護等) を実現

AIP事業実施の考え方（杉山私案）

1. 10年後そしてその先の世界を変え、世界をリードする革新的人工知能基盤技術を構築する。
2. 革新的な基盤技術開発と着実な社会実装の間で、スパイラルに研究開発を推進する。
基盤技術開発と社会実装のダブルトラックにより、10年後およびその先に向けた理論研究を進めながら、その成果をサイエンスや実社会に広く応用する。
3. 当初は緩やかなチーム編成によりやや幅広く取り上げながら、その後、国際的な研究動向、研究進展、成果の見通し等を踏まえて具体的に絞っていくことにより、戦略的な研究開発ポートフォリオを構築する。

革新知能プラットフォーム構築

従来の深層学習を凌駕する革新的なアルゴリズムに基づく人工知能の基盤技術、大規模データベースやハードウェア等の活用に関する統合化技術を一元的に扱うプラットフォームを構築。これにより構築される基盤技術を、順次サイエンスや実社会に応用することで画期的な成果を創出する。

■ 革新的アルゴリズムに基づく基盤技術開発

1. 深層学習等のビッグデータを用いた学習のさらなる抜本的な革新:

現在の技術が適用できないより複雑な構造を持ったデータに対応するための、深層学習等の現在主流の機械学習手法を深化させる革新的超高速学習アルゴリズムを開発する。

2. スパース(疎)・不完全・超高次元のデータからの人間を超える高精度学習の実現:

実社会の実データへの応用で重要となる、偏った情報・不完全な情報・超高次元の情報から人間の専門家を超越するような学習が可能な革新的不完全情報学習アルゴリズムを開発する。

3. 実環境におけるストリーミングデータからの頑健なリアルタイム学習の実現:

多様なセンサーが大規模に用いられる次世代社会で重要となる、次々と与えられるデータを即座に学習し・結果をフィードバックできる革新的リアルタイム学習アルゴリズムを開発する。

■ 人工知能技術の統合化技術開発

1. 学習のための学習: 学習アルゴリズムの選択・調整パラメータのチューニングを自動化。
2. ハードウェアへの最適化: CPU GPUだけでなくDisk I/O等も考慮した新しいデータ処理パラダイムを開拓。

CPU GPU等ハードウェア構成も考慮した超高速並列探索技術の開発。

3. データ収集の最適化: 能動学習・ベイズ最適化などの本質を解明して適用し、データ収集を最適化することで実装における効率化につなげる。

超高齢化社会へ向けた医療サポート

- 動画認識・センサ情報解析・アクチュエータ制御技術を融合し、高度な手術を支援
- 機械学習による電子医療記録の自動解析・予後予測
- 会話を通じて高齢者の認知機能を維持向上

老朽化が進むインフラへの対応

- 自動操縦ドローンを駆使し、動画や打診音から橋梁などの危険個所を自律的に特定

甚大な自然災害への対応

- ビッグデータ解析技術とシミュレーション技術を統合し、甚大災害をもたらす異常気象を適確に予測し、被害を最小限に抑え迅速に復旧できる社会システムを構築

細分化が進む科学研究への対応

- 論文や実験結果等を活用し、科学研究の発展を支援する.
- 超多重検定による科学発見：これまで埋もれていた発見を見逃さない
- 実験計画：能動学習・ベイズ最適化に基づいて、次に実験すべき項目を決定
- スマート・アラートシステム：最新研究開発動向や新たな研究のヒントを示す
- マテリアルズ・インフォマティクス：材料に関する膨大なデータやシミュレーションに基づいて、新材料やその製造法を探索

倫理・プライバシー

- プライバシー，説明責任・透過性，公平性・公正性，差別配慮，セキュリティ

データサイエンティスト、サイバーセキュリティ人材等の育成

- データサイエンティスト教育：クラウドソーシングを用いた遠隔データ解析教育

次世代の人工知能技術の研究開発における3省連携体制

- (1) 各分野でのビッグデータの集積、センサーの量的・質的拡大 (IoT: Internet of Things)。
- (2) 人工知能の50年来の大きな技術的ブレークスルー (自ら特徴を捉え進化する人工知能を視野)。
- (3) 3省連携による研究開発成果を関係省庁にも提供し、政府全体として更なる新産業・イノベーション創出や国際競争力強化を牽引。

