

ビッグデータの利活用のための 専門人材育成について (総括説明)








実質的な
専門知識

平成27年7月30日

大学共同利用機関法人 情報・システム研究機構

ビッグデータの利活用に係る専門人材育成に向けた産学官懇談会

ビッグデータの利活用に係る専門人材育成に向けた産学官懇談会 出席・陪席者一覧

座長	北川 源四郎	情報・システム研究機構 機構長		
委員	安宅 和人	ヤフーCSO/データサイエンティスト協会 理事		
	榎本 剛	文部科学省 研究振興局 参事官(情報担当)		
	岡本 青史	富士通研究所		
	北山 浩士	文部科学省 高等教育局 専門教育課 課長		
	佐藤 俊哉	京都大学医学研究科 教授		
	長谷川 真理子	総合研究大学院大学 理事・副学長(教育担当)		
	樋口 知之	統計数理研究所 所長		
	丸山 宏	統計数理研究所 教授/データサイエンティスト育成ネットワーク事業 実施担当責任者		
	丸山 文宏	富士通研究所		
	渡辺 美智子	慶應義塾大学健康マネジメント研究科 教授/(独)統計センター理事		
陪席	栗辻 康博	文部科学省 研究振興局 数学イノベーションユニット次長/基礎研 究振興課 融合領域研究推進官		
	金井 学	文部科学省 高等教育局 専門教育課 情報教育推進係長		
	栗原 潔	文部科学省 研究振興局 参事官(情報担当)付専門官		
	土生木 茂雄	文部科学省 高等教育局 専門教育課 視学官		
	山路 尚武	文部科学省 高等教育局 専門教育課 課長補佐		

1. データサイエンティスト育成の必要性と我が国の課題


—ビッグデータと情報技術により，従来とは全く異なる価値が提供されるようになった現代—

①ビッグデータの活用領域



異常検出・予測・自動化・最適化

【例】

ネット広告におけるリアルタイムビidding
金融における高頻度証券取引
クラウドソーシングに基づく新しいビジネス
自動運転車  等

③我が国の教育体制・リテラシー状況



1. “統計”を数学，経済，医学，工学等
応用分野に埋め込んだ形で点在
⇒「機会学習」「自然言語処理」「大規模データ処理」
などの最新の技術スキル習得の限界
2. 我が国の人口当たりの理工系の学生数，
アナリスティカルスキル保有者数が極小
⇒ 理系の素養を持つ人が少ないという根本問題
国民全体のリテラシー不足



②現代の我が国のビジネス運営・政策策定








未だ 【KKD（勘と経験と度胸）】

その要因は・・・？

データの利活用に係る教育の不足

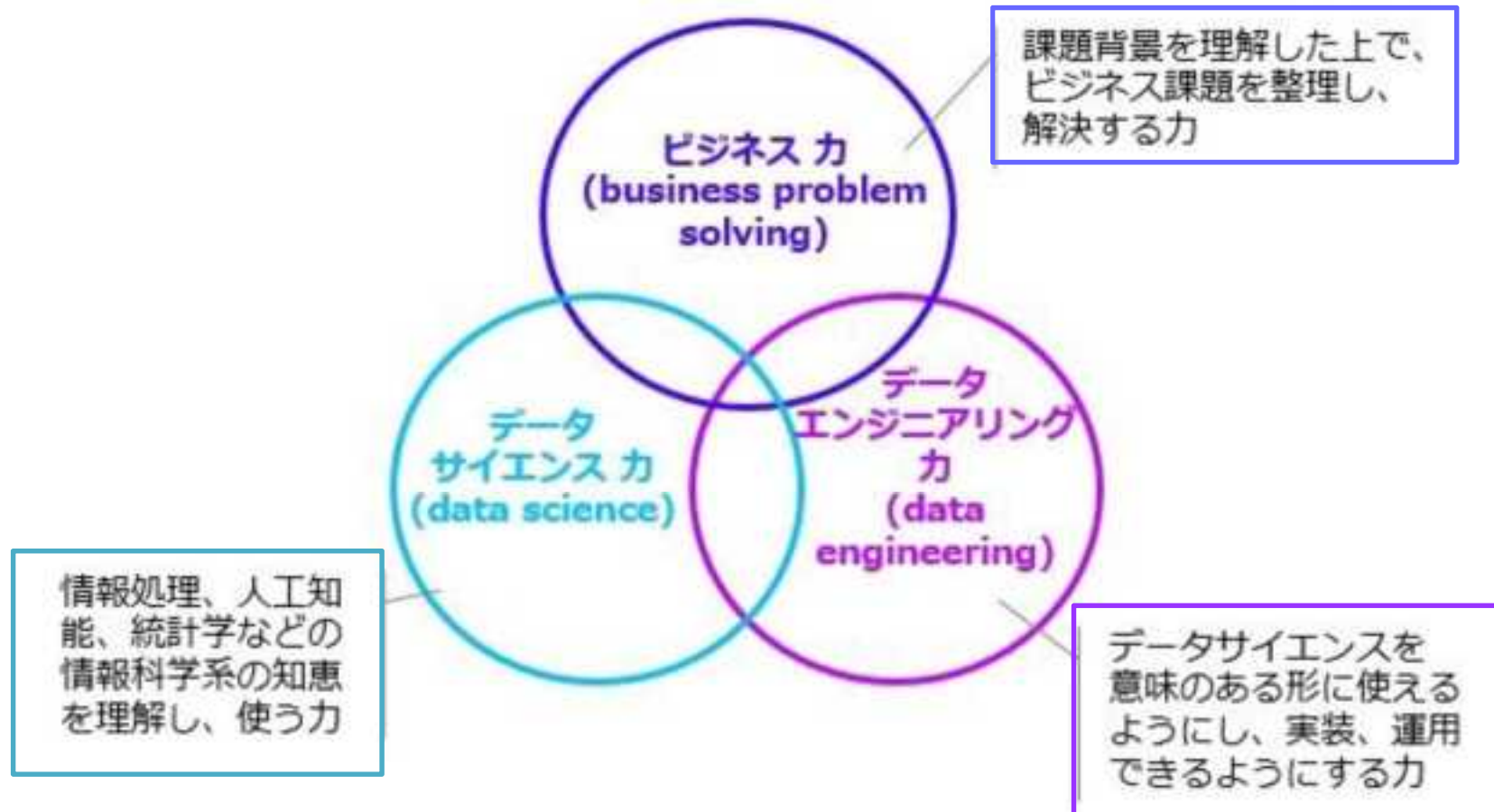
⇒ビッグデータの利活用に係る専門人材
＝「データサイエンティスト」
の育成が急務

現代において求められるものは…

- “伝統的な統計学”の枠を越え
「機械学習」「最適化」「自然言語処理」
「因果推論」など最新の技術を
様々な分野に応用できる専門人材(T型・II型)
- T型・II型人材を実務  
や最先端な研究開発の   
場で活かせる マネージ
メントレベルのデータ
リテラシー

2. 我が国におけるデータサイエンティストへの要請

データサイエンティストに求められるスキルセット



2. 我が国におけるデータサイエンティストへの要請 —産学からの要請—

産業界の要請

- ① 大学でのデータサイエンティスト育成のためのプロフェッショナル教育
- ② 中等・高等教育を含む理系素養・データリテラシーのテコ入れ
- ③ 国家レベルのビッグデータ活用フラッグシップ・プロジェクト

【産業界でデータサイエンティストに求められる能力】

- ① コミュニケーション能力・翻訳能力
⇒顧客と建設的な会話をする・顧客の課題をデータ分析や情報技術に落とし込む
- ② 発見能力
⇒課題領域を見通して本質的な問題＝イシューを見抜く
- ③ 動員力
⇒課題解決のため各分野のエキスパートを動員する

アカデミアの要請

- ① 宇宙・地球物理，高エネルギー物理学物質・材料科学，生命科学，医療などの分野の専門課程を学ぶ学生が，必ずデータサイエンティストとしてのスキルも学べる仕組み
- ② 「第4の科学」＝ビッグデータの分析を行いデータに基づく意思決定を行う方法論を牽引できるトップレベルのデータサイエンティスト研究者
- ③ リテラシーレベルでの教育の一環においてデータに基づく思考ができるようにすること

- ◆ 複数の分野にまたがって議論するところに大きな発見や発明がありうる
- ◆ データサイエンスは複数の分野の研究者が視野を共有するための共通言語である
- ◆ データサイエンティストは分野を結ぶ本質的な媒介者となる

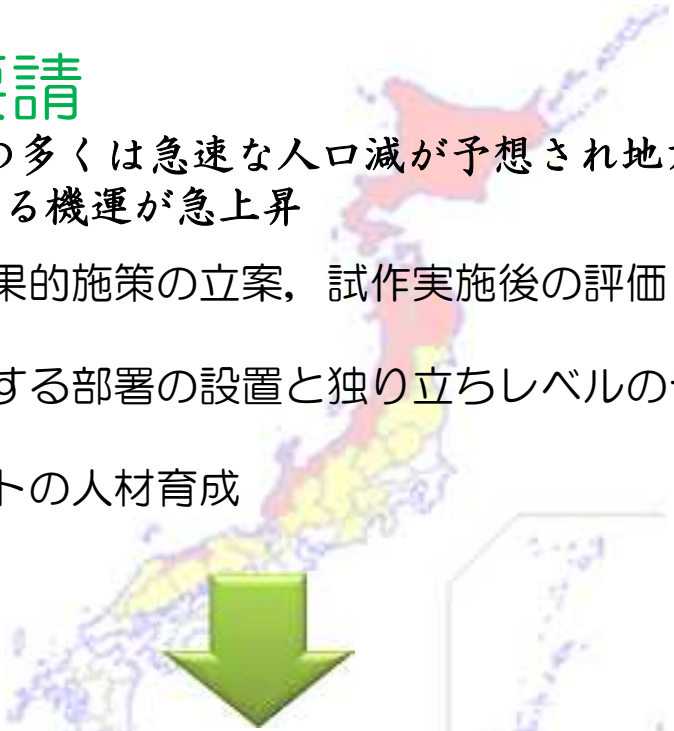
＝イノベーションを牽引＝

2. 我が国におけるデータサイエンティストへの要請 —官からの要請—

地方自治体の要請

我が国約1700の市町村の多くは急速な人口減が予想され地方創成のためデータサイエンティストを求める機運が急上昇

- ① エビデンスに基づく効果的施策の立案，試作実施後の評価・再検討を担う人材の確保
- ② データを戦略的に活用する部署の設置と独り立ちレベルのデータサイエンティスト
- ③ データサイエンティストの人材育成

- 
- ◆ ビッグデータビジネスやデータサイエンティスト人材育成に係るイベントや機会が東京に集中している
⇒地方自治体での人材育成が困難な状況

＝全国レベルの底上げの必要性が明白＝

3. データサイエンス人材育成のあるべき姿と現実に向けた仮設 —育成レベルと、毎年の育成目標人数—

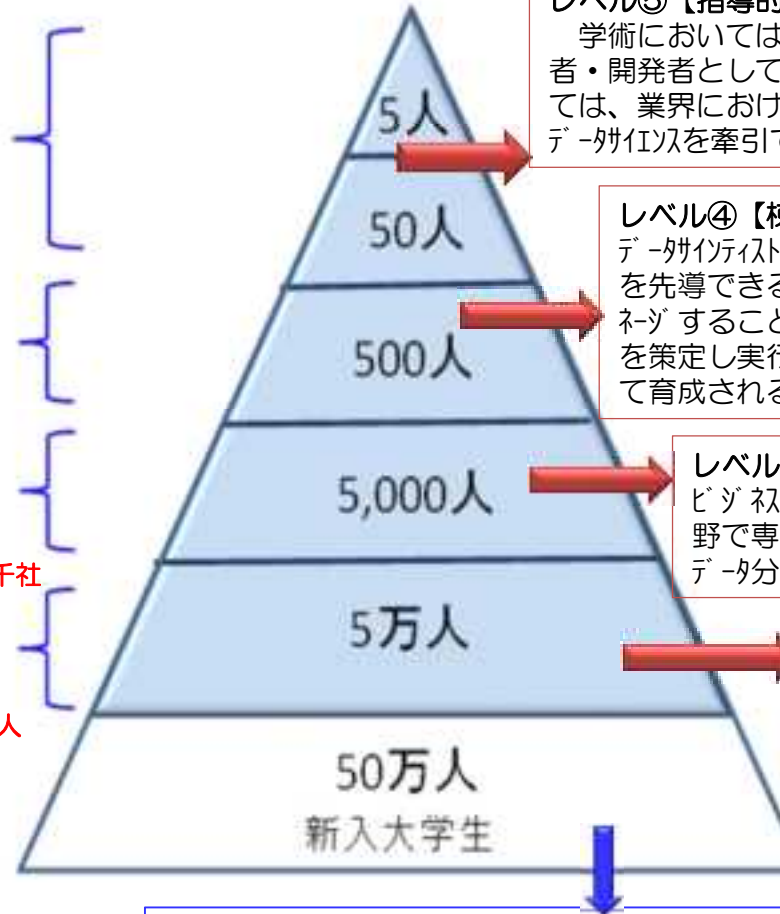
データサイエンティスト協会が定めたスキルレベル (2014年12月)

業界を代表するレベル
Senior Data Scientist
トップタレント
年間数名～数十名

棟梁レベル
(full) Data Scientist
毎年5千人の「独り立ち」を指導統括
「独り立ち」6～15人につき1人程度

独り立ちレベル
Associate Data Scientist
資本金10億以上の会社はおよそ6千社
産業界向きだけでも5千人程度

見習いレベル
Assistant Data Scientist
理系修士入学者は年間およそ5万人



レベル⑤【指導的データサイエンティスト】
学術においてはデータサイエンスの最先端を切り開くワールドクラスの研究者・開発者として指導的な能力を発揮する者、また産業界においては、業界におけるビッグデータ・データサイエンスに基づくビッグデータ・データサイエンスを牽引できるトップタレント (Googleのシリ・ペイブ級の者)

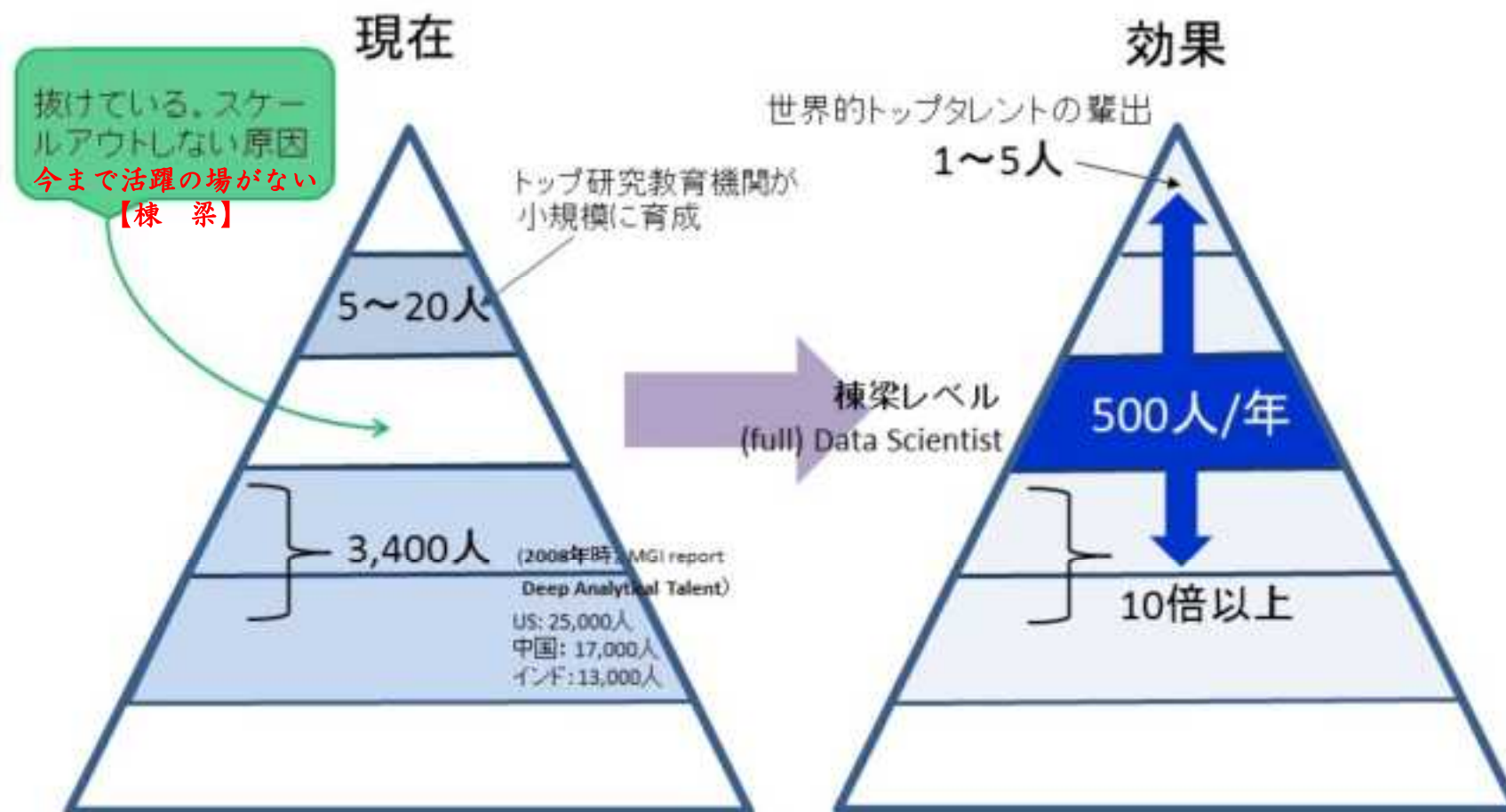
レベル④【棟梁】
データサイエンティストのチームを率いて、組織におけるビッグデータ活用を先導できる能力をもった人。複数の応用分野を俯瞰的にマネージすることができ、データサイエンスの観点から全体最適の戦略を策定し実行するリーダーシップが求められる。主に実務を通して育成される能力

レベル③【独り立ち】
ビジネス、データサイエンス、データサイエンス・アナリティクスのいずれかの分野で専門的な能力を持ち、自らの取り組みで高度なデータ分析・問題解決能力を発揮する。実務経験が必須

レベル②【見習い（基礎能力）】
全てのデータサイエンティスト（実務家・研究者問わず）が持つべき能力。ミドルクラスのマネジメントにも必須。理系の修士は全て、文系でも社会科学系・言語学・心理学等の専攻で身に付けるべき。適切な指導の下、ビッグデータ活用プロジェクトの一部を担当可

レベル①【データリテラシー】
文系・理系を問わず全ての学生が持つべき高校から大学学部レベルの素養（リテラシー）ベースになる統計的概念、データに基づく思考や問題解決の基礎理念、ITリテラシー他

3. データサイエンス人材育成のあるべき姿と現実に向けた仮設 —育成が遅れ、最も欠けている棟梁レベル(左)と育成実現時の波及効果(右)—



- ✓ 棟梁レベルの人材が年**500**人規模で育てば、その中から世界的なトップタレントが現れてこの分野全体を引っ張っていくことが期待できる
- ✓ 同時にこの棟梁レベル人材が、実務の傍ら随時大学や企業で人材育成を支援することで、独り立ちレベル以下の人材育成が促進され、スケールアウトが進むことを期待することができる

4. 具体的施策 ①

① リテラシーの醸成 (50万人規模)

◆ 高校教育・大学教養での講義：

- 高校生・大学生がワクワクするような啓発書・教科書をつくる。
- 社会でどう使われ役立っているかを示す事例のビデオ素材をつくる。
- 活躍中のデータサイエンティストをプールし、大学や高校にこれらのデータサイエンティストを適宜派遣し講義を担当してもらう。

◆ 大学基礎教育：

- 大学124単位の内、共通教育で例えば4単位、専門教育では専門に応じて例えば2から6単位をデータサイエンス（統計）に割り当てる、と定める。この際、核となる週1時間の講義にコンピュータ実習や問題を解く演習もセットにする。
- これに合わせて、基礎統計教育も見直す。

◆ 社会一般の興味を引くために：

- 全国的なデータサイエンスに関するコンテストを実施する。スーパーグローバルハイスクール指定校等の先進的な取り組みを行っている高校にも積極的参加を呼びかける。
- gaccoにおける「社会人のためのデータサイエンス入門」のように、MOOC（Massive Open Online Course）を利用した一般向け教材を充実させる。



4. 具体的施策 ②

② 見習いレベルの育成 (5万人規模)

- ◆ データサイエンス力の **基本スキルの学習** :
= 統計学、機械学習、最適化およびビッグデータ解析に必要なプログラミングやデータ可視化の基礎等

*学部・大学院におけるデータサイエンスの教育カリキュラムは、統計教育参照基準・情報教育参照基準に相当するものを早急に作成

- ◆ **ダブルディグリー・ジョイントディグリー**の考え方@大学院の導入 :
⇒ 専門科目と共にデータサイエンスを副専攻などの形で学習(主副逆も然り)

- ◆ スケールアウトの方策 :
= MOOC (Massive Open Online Course) の積極的利用・コンテンツの充実
データサイエンス向けのコンテンツの充実MOOC 修了者に対して修了書を発行する、あるいは、大学における単位認定を行うなど、MOOC の位置付けの社会的認知向上を図る

➢ 参考取組例 : gacco における「統計学I, II, III」等

- ◆ 社会人対象プログラムの提供 :
特に機械学習など新技術に関して、ミドルマネジメント層を含む広い対象に、**再教育のプログラムを提供**。地方を含む各大学に展開。社会人の学び直しを推進

➢ 参考取組例 :
文科省において検討が進められている「職業実践力育成プログラム」認定制度

(大学等における社会人や企業等のニーズに応じた実践的・専門的なプログラムを国が認定する制度) の活用等



4. 具体的施策 ③

③ 独り立ちレベルの育成 (5千人規模)

◆ 独立してデータサイエンスを推進できるレベル:

問題設定能力、問題解決のための戦略立案能力、データの収集・キュレーション能力、データ分析結果の業務や事業への実装能力、異分野研究者や事業者との連携・コミュニケーション能力、研究倫理、情報セキュリティの能力実装

◆ 重要スキル:

データ解析スキル (=MCMC (マルコフ連鎖モンテカルロ) 法、データ同化、インピュテーション技術、高次元空間の構造探索とモデル化、異種情報統合による個人化技術、隠れた関係の検出、特異性の発見、因果推論等)

データ可視化技術 (=次元圧縮、特徴抽出、パターン認識など)

* 但し、重要スキル等については、データサイエンスを巡る状況の急速な変化に伴う不断の検討が必要

◆ 大学院におけるPBLに基づく専門育成プログラムの推進:

▶ 参考取組例:

「分野・地域を越えた実践的情報教育協働ネットワーク(enPiT)」
「先導的ITスペシャリスト育成推進プログラム」などにおけるPBL

◆ 企業との連携を通じた企業の事例のPBL化、インターンシップ・プログラムを通じた実務学習

▶ 参考取組例:

産学連携では経団連主体のCeFIL (高度情報通信人材育成支援センター)



4. 具体的施策 ④

④ 棟梁レベルの育成 (500人規模)



【再教育 (案)】

情報・システム研究機構
「データ中心科学
リサーチcommons」
統計数理研究所
「統計思考院」
で行われている、データ
サイエンス研究者の育成
プログラムを量的に拡充

アカデミア及び産業界で
活躍できる棟梁レベルの
データサイエンティスト
を育成

◆ 国家的な拠点設置の必要性：

⇒ビッグデータ・データサイエンスの最先端の手法・応用の研究・開発及びそれらに精通した人材育成を行うための国家的な拠点設置が必要

◆ 棟梁レベルデータサイエンティスト育成の集中的プログラム開講：

⇒データサイエンティストとしての実務経験のある社会人を、棟梁レベルに育成する集中的プログラムを上記拠点で開講

*分野を俯瞰し戦略の立案・実行ができる棟梁レベルの人材は、常にデータサイエンスの最新的手法群(ノンパラメトリックベイズ、スパースモデリング、カーネル法、グラフィカルモデル、深層学習、分散リアルタイム計算フレームワークStorm13、広告におけるリアルタイムビディング、マーケティングにおける協調フィルタリング、ロボティクスにおける画像・映像認識、質問応答システムの医療応用など)への精通が必要

*最先端の手法をPBLを通して実地で経験し、各応用領域での最新の成果をケーススタディとして学ぶと共に、拠点の特質を活かして棟梁レベルデータサイエンティスト間の人的ネットワークを形成

▶ 参考取組例：情報学研究所「トップエスイープログラム」
統計数理研究所「データサイエンス・リサーチプラザ」

◆ 活躍できるキャリアパスを見据えた再教育

⇒科学の諸分野において、既にビッグデータを活用した研究方法論をある程度身につけた人材(博士号取得者等)に対して、産業界やアカデミアにおいて活躍させるための再教育

▶ 参考取組例：米国Insight Data Science Fellow Program

4. 具体的施策 ⑤

⑤ 指導的データサイエンティストの育成 (50人規模)

- ◆ ワールドクラスの人材を系統的に育成するのは困難
⇒ 世界最先端の手法・応用の研究・開発を推進することによって、このような指導的データサイエンティストが生まれてくる土壌を醸成することが重要
 - 国レベルで、ビッグデータのフラッグシップ・プロジェクトを実施する中で指導的データサイエンティストが活躍できる場を提供
- ◆ 才能のポテンシャルを持つ若手を早い段階で発掘し十分な機会を提供
 - 定期的なデータサイエンス・ハッカソンを実施
 - *IT セキュリティの分野で、同様の試みとしてIPA「セキュリティ・キャンプ」が存在
 - これを参考に、データサイエンスにおけるプログラムを企画
 - メンター制度・資金援助機会の提供
 - *産業界で活躍する指導的データサイエンティストになれる人材を育成するために、才能のある若い者にメンターをつけ、資金等の援助を与え人脈を形成する機会を提供。
 - *IPA「未踏IT 人材発掘・育成事業」やJST「さきがけ」を参考に、指導的データサイエンティスト候補者向けのプログラムを作成。



5. まとめ

【ビッグデータ利活用専門人材の出口戦略】

- 育成された人材は、民間・アカデミアを問わず広く活用されなければならない
- ビッグデータ利活用とデータサイエンティストの重要性を広く社会に発信していくことが重要
- データサイエンティスト教育を受けた者、あるいはある一定のスキル基準を満たす者に対してスキル認定を行い、雇用する側とデータサイエンティストのスキルのミスマッチが起こらないようにすることも重要

【ビッグデータ利活用専門人材育成の課題と解決に向けた取り組み提案】

1. 人材不足問題の根源を解決するために
⇒「棟梁レベル」の決定的不足を解決のために国家レベルの拠点を設置
年500名規模の「棟梁レベル」の人材育成をめざし、上層への成長や下層へのトリクルダウン効果も狙う。
2. 「リテラシーレベル」「独り立ちレベル」の教育加速のために
⇒リテラシーレベルや独り立ちレベルの大学教育を加速させるために、主要10大学程度で本報告書の提案に基づく人材育成をスタートすると共に、MOOCなどのオンライン教材を整備し、全国への波及効果を狙う。
3. 社会全体のリテラシー・アウェアネスの向上のために
⇒全学的教養教育の実施、国家レベルのフラッグシップ・プロジェクトの推進、コンテストの開催、映像素材の充実などの取組を行う。

懇談会報告要旨 (1 / 2)

1. 報告書の背景

情報通信技術の発展による社会の価値がビッグデータとそれを処理する情報技術へ大きくシフト
我が国ではデータサイエンティストは大幅に不足
データサイエンティスト育成体制なし

⇒データサイエンティスト育成が喫緊の課題

- I. 産業界・アカデミアからの要請の把握
- II. 育成すべきデータサイエンティストが持つべきスキルや能力の特定
- III. データサイエンティスト育成のための方策

を懇談会で検討

2. データサイエンティストのスキルレベルとその育成

データサイエンティストのスキルレベル

- データサイエンス力（統計学、機械学習、最適化などを理解し、使える力）
- データエンジニアリング（データサイエンスを実装する力）
- ビジネス力（課題を理解し、問題設定し、解決する力）

など



データサイエンティスト育成を、すべての大学生を想定
リテラシーレベルから世界トップレベルまで6つのレベル
で考えることとし、その育成方法を懇談会で検討

懇談会報告要旨（2 / 2）

3. 提言

- ◆ 我が国の問題の根源は、**棟梁レベルの決定的不足**にある。この解決のために**国家レベルの拠点を設置**して、年500名規模の「棟梁レベル」の人材育成をめざし、上層への成長や下層へのトリクルダウン効果も狙う。
- ◆ リテラシーレベルや独り立ちレベルの**大学教育を加速させる**ために、主要10大学程度で本報告書の提案に基づく人材育成をスタートすると共に、MOOCなどのオンライン教材を整備し、全国への波及効果を狙う。
- ◆ **社会全体のリテラシーやアウェアネスを向上させる**ために、全学的教養教育の実施、国家レベルのフラッグシップ・プロジェクトの推進、コンテストの開催、映像素材の充実などの取組を行う。



これらの方策を実現するにあたっては、データサイエンスを副専攻とする**ダブルディグリー制****人材育成の推進**や**スキル認定制度も有効**と考えられる。