

座談会

ビッグデータ，IoT，AI：最新の事例と人材育成

進行役：石井一夫（久留米大学）

インタビュイー：里 洋平（DATUM STUDIO），橋本武彦（（株）GA technologies），
福中公輔（（株）GA technologies），加藤 浩（放送大学）

インタビュア：吉野松樹（（株）日立製作所）



石井一夫（久留米大学）

久留米大学バイオ統計センター准教授，1995年徳島大学大学院医学研究科博士課程修了。博士（医学）。専門はビッグデータ分析，計算機統計学。R，Python，Juliaなどのオープンソースソフトウェアを使用して医療ビッグデータ，医療ゲノムデータ，医療IoTデータを用いた大規模データ分析に従事。



里 洋平（DATUM STUDIO（株））

R言語の東京コミュニティTokyo.R創立者。ヤフー（株）で，推薦ロジックや株価の予測モデル構築など分析業務を経て，（株）ディー・エヌ・エーで大規模データマイニングやマーケティング分析業務に従事。その後，（株）ドリコムにて，データ分析環境の構築やソーシャルゲーム，メディア，広告のデータ分析業を経て，DATUM STUDIO（株）を設立。



橋本武彦（（株）GA technologies）

（株）GA technologies, AI Strategy Center. 国立大学法人電気通信大学 客員准教授, 帝京大学文学部社会学科卒. 専門はデータ分析, PropTech（不動産テック）をはじめとしたReal×TechのX-Tech領域におけるAI・データ活用の啓発やデータサイエンティスト・AI人材の育成に従事.



福中公輔（（株）GA technologies）

（株）GA technologies, AI Strategy Center. 早稲田大学大学院文学研究科人文科学専攻博士後期課程単位取得後退学. 博士（文学）. 専門は統計学, データ解析, テスト理論. 不動産や金融関連データを分析し, AIを活用した新規サービス立案とビジネスの効率化等に従事.



加藤 浩（放送大学）

放送大学教養学部教授, 熊本大学客員教授, 北海道大学客員教授. 1999年東京工業大学大学院社会理工学研究科人間行動システム専攻博士課程修了, 博士（工学）. 専門は教育工学, 特に社会文化的アプローチより協調学習を支援する仕組みやシステムの研究開発に従事.



吉野松樹（（株）日立製作所）

（株）日立製作所、IoT・クラウド事業部データマネジメント本部シニアプロジェクトマネージャ、本会論文誌トランザクションデジタルプラクティス編集委員長、資格制度運営委員会副委員長、本会フェロー、博士（情報科学）。

はじめに—本日の出席者—

石井：それでは、論文誌デジタルプラクティス（以後、本誌）、ビッグデータ特集号の座談会を始めたいと思います。私は久留米大学バイオ統計センター准教授の石井一夫と申します。本日の進行役を務めさせていただきます。今回は、「ビッグデータ、IoT、AI：最新の事例と人材育成」というテーマで最新のビッグデータ、IoT、AIをめぐるトレンドや人材育成の動向をお話いただき、今後のこの分野の在り方、展望などを見ていければと考えています。

今回、本誌でビッグデータ特集号を企画するきっかけになったのは、昨年のソフトウェアジャパン2019でのビッグデータセッションを企画したときに、本日の座談会に出席されている当時本誌の編集長をされていた吉野松樹さんにお声がけいただいたことからです。ちょうど、「ビッグデータ解析のビジネス実務利活用（PBD）研究グループ（以後、ビッグデータ研究グループ）」という研究グループの立ち上げをしていたときで、今日いらっしゃっている、DATUM STUDIOの里洋平さん、（株）GA technologiesの橋本武彦さん、福中公輔さんは、ビッグデータ研究グループの立ち上げや、ソフトウェアジャパンの企画にご協力いただいていた関係でお声がけしました。放送大学の加藤浩先生は、（私、石井も委員として参加しているのですが）本会のデータサイエンス教育委員会の委員長をされていて、現在「大学のリテラシーレベルの数理・データサイエンス・AI教育」に関する議論を展開しています。その最新動向をお話いただけることを期待しております。なお、今回の本誌ビッグデータ特集号では、私、石井が「医療ビッグデータアナリティクスプロセス—抗がん剤副作用の解明における実践—」と、福中さんと橋本さんが「情報の非対称性の解消に向けた中古マンション価格推定の取り組み」という論文を寄稿しており、これに関連したコメントも織り混ぜていくことも予定しています。

第1部：ビッグデータやデータサイエンスとのかかわりを絡めた出席者の自己紹介

1.1 進行役（石井一夫（久留米大学））：医療ビッグデータとビッグデータ研究グループ

石井：まず、最初に本日まで参加していただいている方々、一人ひとりにビッグデータとか、データサイエンティストを絡めて、自己紹介をしていただきたいと思います。

まず、進行役である私から最初に自己紹介させていただきます。私は久留米大学バイオ統計センターで准教授として、主に医療に関するデータモデリングや、データ分析の教育、研究に携わっています。

今の主な研究テーマは医療ビッグデータで、医療電子ドキュメント（EMR）であるレセプトデータの分析とか、次世代シーケンサーデータを用いた医療ゲノムデータの分析とか、AIを用いた診断画像の分析などもやっています。ですから、医療に関するビッグデータ、データサイエンスに関する全般的なことをやっているということになります。

久留米大学バイオ統計センターでは、ほかの先生方は、主に統計分野にフォーカスしている一方で、私自身は機械学習とか、自然言語処理とか、画像分析とかいろいろ手掛けていますので、他の統計専門の教員スタッフよりは専門が少し違っていますが、いろいろな方面で尽力しています。元々は、20年くらい前にゲノム解析からデータ分析の世界に入ってきました。

本会では、2013年6月にITフォーラム「ビッグデータ活用実務フォーラム」というコミュニティを設立し、以来、本会ソフトウェアジャパンのビッグデータセッションや、論文誌デジタルプラクティスのビッグデータ特集号を企画するなどの活動を行ってきました。その後、「ビッグデータ活用実務フォーラム」から発展して、2018年度に「ビッグデータ研究グループ」を設立して活動を展開しており、ビッグデータに関するアカデミックな基盤を構築できないかと考えております。

本日はよろしくお願いします。

1.2 インタビュイー（里 洋平（DATUM STUDIO））：ビジネスにおけるデータサイエンスの展開

石井：それでは、里さんから、自己紹介をお願いします。

里：DATUM STUDIOという会社の取締役副社長CAOをやっている里と申します。

自己紹介として、ビッグデータや、データサイエンスを絡めてお話をさせていただくと、最初のかかわりは新卒でヤフーという会社に入社したのが始まりになります。2008年入社なので12年ぐらいずっとこの領域でやっていることになります。まだ当時はビッグデータとか、データサイエンスという言葉はそんなに言われていなかった時期ですが、ご存知の通りヤフーというのは大きい会社で検索エンジンも持っていて、大量のデータを持っている、まさにビッグデータという領域かなと思います。

その中で、動画のレコメンドエンジンを作ったりしてました。YouTubeを見ていると、隣に関連動画とかがあって、放っておくと次にその動画が動き出す。まさにああいうのをやっていた。

ほかには株価の予測みたいなものをちょっとやっていた。まあそんなのは結構みんなやっているじゃないかみたいに感じると思うのですが、特別なのは、Webの記事やSNSの情報を使って、ある企業に対してポジティブに書かれているのか、ネガティブに書かれているのかという

のを抽出し、それをスコア化するところでした。その企業のインターネット上での評判ともいえるスコアをつくって、それと株価の動きをモデリングするみたいなことをやったりしていました。

ネットとかでよく言われ、記事になりやすい企業は、それなりに予測できたんですが、ゼネコンとかそのあたりは、全然情報がなくて（笑）。すごい限定されたところでした。

その後、ソーシャルゲームがすごい流行った時期にDeNAという会社に移りました。ちょうどその頃、ビッグデータという言葉が出てきた時期で、大規模データマイニングとか、Hadoopを使った分散処理とかが言われてました。DeNAでは、Mobageプラットフォーム上のデータマイニング、ユーザの分析というのや、ゲームのレコメンドや不正ユーザの検知などをやっていました。

後半はマーケティング部の方に移りました。それまでは、結構エンジニアリングの世界にずっといたのですが、マーケティング部門という、ビジネスのほうの世界に移って、そこの中でマーケティングの分析や、テレビCMとかを含めた広告効果分析などをやってました。

その後、ドリコムというソーシャルゲームの会社に移りまして、そこでは主にデータ分析組織の立ち上げをやりました。もう少し具体的に言うと、分析者、今だとはデータサイエンティストと呼ばれる人たちの生産性を上げるというのをミッションにして、そのためのデータ分析の環境の構築や人材育成などです。そのほかにも新規事業に対してデータをどう活用するのかの支援もやっていて、ログ設計をしたり、シミュレータをつくったりということをやっていました。

その後、今のDATUM STUDIOという会社を立ち上げまして、データ分析や機械学、人工知能といったもののビジネス活用を支援することをやっています。

立ち上げた当初はやはりWeb系とか、ソーシャルゲーム系のお客さんが多かったんですが、今では、Webとか、ソーシャルゲームの案件はほとんどなく、それ以外の業界がほとんどという状況です。ここ5～6年ぐらいですごく景色が変わったなというふうに感じております。

ということで、今日はわりと自分の会社から見たデータサイエンスとか、そこら辺の話、実情みたいなのができればなと思っております。よろしくお願いします。

1.3 インタビュイー（橋本 武彦（（株）GA technologies））：データサイエンスへのかかわりと、人材育成、データサイエンティスト協会

石井：次に橋本さんお願いします。

橋本：不動産の事業会社である（株）GA technologiesの橋本と申します。電通大には、文科省のAI人材育成の活動でD-DRIVEというのがあり、電通大、阪大など、今、5拠点ぐらいあります。電通大ではデータアントレプレナーという言い方をしているのですが、その支援をずっとやらせてもらっていて、その客員准教授も務めております。

ビッグデータ、データサイエンスのかかわりを申し上げますと、元々ブレインパッドというデータ分析、データサイエンスの会社に10年ほどおりまして、そこで里さんと同じで、クライアント向けのデータ分析をずっとやっておりました。その中で里さんとは、TokyoWebminingやデ

ータサイエンティスト養成読本なんかで一緒にさせてもらったご縁があります。

あと石井先生とのかかわりで申し上げますと、さっきお話が出たデータサイエンティスト協会というのが前職のブレインパッドが発起人となって立ち上げてまして、それが2013年の話なのですが、たまたま私がそのときの立ち上げメンバになっていまして、事務局長を一昨年まで担当しておりました。

あとは、前職の後半は、データサイエンティスト育成の新規事業の立ち上げを担っておりまして、そのご縁でアカデミアの先生ともご縁ができ、放送大学で言うと「身近な統計」の渡辺美智子先生にはいろいろお世話になりました。

あと、僕は知らなかったのですが、日立製作所さんも最近データサイエンティスト協会のメンバに入っていただきまして、発足時からずっとお願いしていたのですが、私が事務局長を退任してから入ってくれたようでして（笑）、そんな感じで、いろいろなところに顔を出しています。よろしくお願いいたします。

1.4 インタビュー（福中 公輔（株）GA technologies）：統計学の理論研究者、データ解析コンサルタント、事業会社データサイエンティストへと展開

石井：次に福中さんお願いします。

福中：橋本と同じくGA technologiesの福中公輔と申します。今やっている仕事というのが不動産に関連するデータの分析を、プロジェクトマネージャ的な立ち位置で推進していくというような仕事です。僕は元々早稲田大学で、統計学者の豊田秀樹先生に師事しておりまして、その豊田先生のところで博士号を取ったという感じです。それで、元々はその早稲田の方で助手をやって、統計学の理論研究者としてずっとキャリアを積んできたのですが、まだそれこそ里さんがおっしゃっていたようにビッグデータとか、そういう言葉がなかった時代に、僕の中でこれから企業で恐らくこういうデータ解析系のブームがくるのではないかという思いがあって、そのデータ解析コンサルができるような組織（産業能率大学総合研究所）に転職をしたというような経緯になっています。

それでそこで東京メトロのデータ解析支援をやらせていただいたりもしながら、企業のデータ分析をずっとやってきました。それで、そのデータサイエンティスト協会が主催しているデータサイエンスアワード2017で、東京メトロと共同で発表して最優秀賞を受賞いたしました。

産能大でそのままデータ解析コンサルみたいなことを続けていくこともできたのですが、やはり相手のデータを使って何かをするというところにちょっと限界を感じていました。それよりも自社のデータを使った方がもっと大きなことができるのではないかというような思いもあったところに、ちょうど橋本から声をかけてもらってGA technologiesへ転職という感じで今があります。

僕のデータサイエンスとのかかわりというか、キャリアはそんな感じで、あと専門は元々は構造方程式モデリングをメインでやっておりました。どうぞよろしくお願いいたします。

1.5 インタビュー（加藤 浩（放送大学）：本会のデータサイエンス教育委員会とデータサイエンス教育モデルカリキュラム

石井：次に加藤先生お願いします。

加藤：放送大学の加藤と申します。私自身はずっと教育工学をやっているんで、教育畑なので、だからデータサイエンスとのかかわりと言っても、教育ですから、心理的な実験なんかもするので、統計学はもちろん使いますし、量的な分析も、質的な分析も、両方とも行うというぐらいいのことでした。AIとかはあまり関係してこなかったのですけれども、本会のデータサイエンス教育委員会の委員長を拝命することになりまして、そういう関係で今その大学のデータサイエンスのモデルカリキュラムのまわりの仕事をさせていただいております。簡単ではございますけれども、そんなところです。

1.6 インタビュー（吉野松樹（日立製作所））： 本会論文誌デジタルプラクティス特集号のコーディネータとして

石井：最後に吉野さんお願いします。

吉野：日立製作所の吉野です。データサイエンスという観点で言うと、私は自分で分析をする立場ではありません。現在の所属は、データマネジメント本部ですが、データを貯めるためのデータベースの担当で、皆さんの方から見ると、縁の下の力持ちみたいな、ソフトウェアを開発している部署にいます。

今日なぜここにいるかというと、デジタルプラクティスの編集委員長を2019年3月末まで務めていました。1月15日に発行されたデジタルプラクティスの最新号では、DX時代の人材育成という特集を組みました。また、この特集号のコーディネータとして石井先生に企画いただいた特集の企画がうまくいくように、著者の方をフォローしたりだとか、著者の方についている編集担当者の方に早くコメントを返してくれとか（笑）、そういうことをやっているという立場で参加しています。

あと、本会の資格制度運営委員会の副委員長もやっていて、そこではデータサイエンティスト資格というのはどうあるべきかみたいな議論を始めています。ということで、少し素人的な観点から今日はお話を聞かせていただければなと思っています。よろしくお願いします。

第2部：ビッグデータ， AI， IoTの今のトレンドと将来展望

2.1 医療におけるビッグデータのトレンドと今後の展開

石井：それでは、出席者の皆さんに、職場との関係とか、皆さん方のお仕事と絡めて、この業界の今のトレンドなどを、お話いただいて、その後、その将来展望を話していただきたいと思っています。

進行役として私が最初に、医療ビッグデータの現状について、お話をさせていただきます。医療では、データは毎日のように大量に生産されています。レセプトデータですとか、健康診断のデータですとか、電子カルテとか、それがデータベースに大量に蓄積されています。国家レベルでも厚生労働省が大体170億件とかのデータを蓄積してしまっていて、そういったデータを解析するというようなことをやっています。いわゆる医療ビッグデータ分析というのが1つのトレンドになっています。

当然のことながら百数十億件のデータを一挙に扱えるという人材は希少なのですが、私は今その真っ只中にいて、この分野に入る前にいた大学では、ゲノム解析ということをメインにやっていたので、そのゲノム解析のテクニックがそのまま使えるということなので、それをうまく活用させていただいています。

それで医療ビッグデータということで、AIなどゲノムなどを含めているいろいろなデータ分析を行っている、ということは、先ほどお話しした通りなのですが、今後の流れということ言えば、このビッグデータの流れは止まらないというか、ビッグデータ分析とか、医療のITの活用促進とか、どんどん進んでいくであろうと思っています。医療では、非常に深刻な問題を抱えています。それは少子高齢化で、高齢者の人口が、今後どんどん増えていって、2050年か、2060年頃には、高齢者1.3人を1人の労働者で養っていくというような、ほとんど1人の労働者が1人の高齢者を養わないといけないような人口のアンバランスというのが非常に増していったら、その中で医療システムをどうやってきちんと運営していくか、きちんと皆さんに医療のサービスを提供するかというようなことが将来的な問題となっています。そのあたりをAI—ビッグデータを活用して、どうやって解決策を見出していくかが課題となっています。

2.2 データサイエンスビジネスのトレンドと今後の展開

石井：次に里さんに、現在のご自身のビジネスの展開で、今のトレンドとか、将来的な展望とかの話をお願いします。

2.2.1 データサイエンスビジネスで求められる人材の変化

里：我々から見ている景色というのは結構やはりビジネスの世界の方が多いのですが、今、先進企業の中では、データの利活用を進めて、いわゆるPoC、実証試験というところのフェーズをもう終えて、システム導入の時期に入ってきているなというふうに見えます。これまでは実証実験というところでその企業さんが持っているデータであったりとか、ビジネスだったり、機械学習って本当に使えるんだっけ、うまくいくんだっけ、データ分析にデータはこれで足りるんだっけとかということを、検証するというフェーズがやはり多かった。なので、数年前、3年ぐらい前だとまだまだシステム化という話はあまりなかったです。できるかどうか分からない状況なので、システム化を想定して何かいろいろやろうというのはかなりリスクです。それに投資する価値があるのかどうかという検証フェーズでしたので、求められるのはいわゆるデータ分析とか、機械学習のスキルでした。けれどもいまはシステムの導入時期に入っているのだから、機械学習のモデルだけではなくて、それを実際に運用するときに耐えるそのシステム化のところですね、システムを組めるエンジニアリングスキルというのもすごい求められていて、それがなくて逆にもう価値を出せなくなっている。というのも機械学習というのは今いろんなツールが出てきている、ある程度自動化というのもできるようになってきている。

2.2.2 データサイエンスビジネスにおける人材育成の変化

里：教育においては、教材もWebや、本、動画など大量に出ているところもあって、その部分だけでできる人はもう結構わりといるんですよね。データサイエンティストは不足していると言われるけれども、実は機械学習のスキルだけ見るとできる人は結構大量にいるなという印象も

あります。だからそこだけのスキルでやっていくというのは今辛くなってきているところです。実際に運用するシステムが組めるというところまでいって、ようやくちゃんと価値が出せるというフェーズに移ってきているので、我々としても結構踏ん張りどころというか（笑）。

これまで求められていたのはその機械学習のスキルだったので、システムのエンジニアリングのところはちょっと弱くなってしまっている。そこは今後しっかりとやっていかなければいけないという課題感を持っています。

2.2.4 データサイエンスビジネスにおける顧客ニーズの変化

里：そういったニーズというか、システム化という話がある一方で、従来からやられていた、施策の検討だったり、その効果測定みたいなデータ分析というものも、だんだんみんながデータを使うようになってきたこともあって、すごくニーズが高まっているなと思います。こちらに関しては、高度な機械学習のスキルがそんなに求められるわけではないものの、いわゆるビッグデータを正しく効率的にハンドリングしてちゃんと集計するスキルや、ビジネス的に価値への示唆を得るためのビジネスドメインの知識だったり、解釈力みたいなもの、そういうのを持つ人というのが求められているなという印象です。という感じで、二軸ですかね、自動化というところと、ビッグデータをちゃんとさばける、示唆を出せる人が今後求められてくるのかなと思っています。

2.2.4 データサイエンスビジネスにこれから求められる人材像

石井：ドメイン知識とか、プログラミングとか、全部のスキルを一人の人間がカバーするというのは結構難しいと思うのですが、そのあたりはいかがですか。

里：そうですね。実際にその通りで、チームでできるといいなとは思っています。その課題感もあると思っていて、やはり全然違うスキルなのですよ、データをさばくというのと、データを解釈するという知識と、ビジネス的に価値を出すという、いわゆるデータサイエンススキルセットの3つのスキルというのが全然違うスキルなので、難しいのです。逆に言うと、それぞれのスペシャリストはいるのですよね、なので、特にこれからやっていこうみたいと思う人はどこかの1つのスペシャリストになろうとすると結構しんどい。今から自分で価値を出すには、複数のスキルを合わせ持った方が良いでしょう。そういう人だと、チームワークがうまくできるようになる。それぞれのスキルの方向に、ある人はあっち向いて、別の人はこっち向いて、とバラバラな人だけ集まっても、実は勝負にならなくて（笑）。データをハンドリングできる人はちゃんと解釈する人と一緒に、どういうハンドリングをするのか、集計するのかというのをちゃんと決めてやってみて、なんかちょっと違うねと、やり直しのときとかもしっかりとお互い協力し合っているものでやっていくのがよい。よく見かけるのは、「言われた通りに集計したのにまたやり直しだと（怒）」みたいな話で、すごい溝が深くなるケースが多いです。互いの理解が足りないからそうになってしまうので、その部分をうまくできるような、教育というか、人材育成を心がけています。それぞれちゃんと敬意を持ったかたちでやらなければいけないんだというのを、そういうマインド面とか、育成のときに大事なのではないかなと思っていますね。

橋本：言葉が通じないというわけでもなく、お互いに通じているつもりなのだけれども、実際はもう一步とどいていないというところですね。

里：そうですね。うちの会社がまだ小さいときは逆の役割をやらせていました。プレゼンが得意な人にプログラミング、プログラミングが得意な人にプレゼン、お互いに実際に業務を組んでやってみる、そうするとお互いちょっと困る（笑）。結構、アグレッシブですけども、そこまでいくと、尊敬し合うようになる。この人はここができるというのは素晴らしいみたいな。言葉だけだとどうしても、相手の業務を無意識に軽んじてしまうというのが、見えている景色ですね。

福中：AIなどのツールがどんどん進化して、その機械学習などの部分ではあまり差別化ができなくなってきているということですか。

里：そうですね。そのもの自体ではもうちょっと難しいですね。

2.2.5 データサイエンスビジネスに求められる2つのフェーズ：課題設定と前処理

福中：そうすると、次に差別化するとすればどこになるのですか。

里：さっきの求められているものでいくと、2つあって、最初のフェーズと最後のフェーズ。さっき話したのは最後のフェーズで、自動化のシステム化のところですよ。もう1つは最初の課題設定するところで、AIに何を分からせるのかというところの部分をつくれるかどうかというのが差別化になります。

福中：なるほど、そうすると今後はビジネス力の方が重要になってくるということですか。

里：その部分は両方ですね。もうシステム化まで、導入のところまで進んでいるところからするともう課題設定はできてるので、そこに対して、課題を見直しましょうみたいな形にはなかなかやらないと思います。今から取り組みをはじめるというフェーズでは、やはり課題設定のところで価値を出さないといけないかなと思います。あとは地味に、データをつくるところですかね。課題設定して、その課題を解くためのデータセットをつくる場所の部分も、また重要なところなんですけど、そこはまだ自動化できていない。そこが重要なわりにちょっと軽視されているところがあるので、もうちょっと啓蒙した方がいいのかなとは思っていますね。

福中：やはりそういうところは結構いろいろなノウハウがあってという感じなのですか。

里：そうですね。ノウハウもありますし、純粋にでっかいデータとかだと、普通にさばくだけでもスキルが必要です。なので、そこら辺のところできっかりと元の生データから今回の課題解決のためのデータセットというのをつくっていく。さらに、自動化までいくと、それもやらなければいけない。データセットの作成というのが、もしかすると一番重要になってくるかもしれないですね。

橋本：前処理とか言われるような。

里：そうです。前処理。

橋本：そこが8割だとか、9割だとか（笑）、諸説ありますけれども。

石井：医療でも同じですね。数十億件のデータを渡されて、ひと通りデータの構造を解析するところから始めて、そういうのを解析して、それからデータセットをつくるというような作業に入って.....

2.3 DX領域におけるデータサイエンスの動向

石井：次に進めます。橋本さん、福中さんに、不動産関連のデータサイエンスビジネスのトレンドや展望をお話しいただけますでしょうか。

2.3.1 DX領域で求められるデータサイエンスの動向

橋本：今の里さんのお話に全部賛成です。エンジニアリング力と滋賀大（元大阪ガス）の河本さんがおっしゃっている課題を見つける力、解く力、使わせる力のうち、を見つける力と使わせる力での差別化とか、私も本当に賛成です。ここで僕がコメントしようと思ったのは、データやAIの民主化みたいな話です。解く力の部分がツールなどでどんどん進化、自動化して、差別化できるのはその上流か下流かと思います。その中で私は不動産業界に入ったのですが、不動産というのが非常にIT化の遅れた業界で、今で言うクロステックとか、最近だとDXという言い方をしますが、ECとかソーシャルゲームなどがデジタルの先端だとすると、紙とFAXの遅れた業界なのですが、そこではやはりまだテーマ設定みたいなところ、最近だとAIコーディネータという言い方をしますが、要は、何のテーマでどうやって問題解決するかみたいなところが一番ウエイトが大きいのかなと思っています。

別に不動産に限らずですけども、IT化の遅れたクロステックの業界でうまくいっている会社と、うまくいっていない会社の違いは機械学習の腕や知識の差ではなくて、何のテーマをどう解くかというのを、ビジネス現場と一緒にやっていくかの差ではないかなと私は思っています。

あとそこに加えて言うならば、経営層の支え、データ、インフラですかね。ちなみに、私が入社初年度にやった仕事の1つは、会社設立後のすべて紙の契約書を再入力してデータ化・整備するPJでした。

おかげさまで所属のAI Strategy Centerは社外からも評価いただくことが増えていますが、振り返ってみるとテーマ設定とエンジニアリング力によるラピッドプロトタイピングが大きいかと。前職、私はクライアントサービスの会社だったので中間報告までこのモデルをつくってみたいな世界だったのですが、今は、思いついたらクイックにつくれるようになっていて、たとえば営業の人が同じフロアにいますので、これでどうか見せて、フィードバックをいただき、またすぐ直してみたいな感じでやっています。そういうのが比較的やりやすいのが事業会社の優位性なのかなと思ったりします。

2.3.2 DX領域で求められるデータサイエンス人材の動向

福中：僕もまったく同意見で、最初に里さんが言われていたことがほぼすべてだと思っていて（笑）、データ分析のスキルそのものを持たれている方というのは、もう結構世の中にはいるのですよね。僕がやはり今の会社にきて一番驚いたのが、新卒上がりの人でもかなりの分析スキルはもうお持ちなのです。なので、もう分析スキルそのものに関しては最初からかなり高いレベルにあって、それでそういう人たちが、ではビジネスをうまくできるのかというと、やはり就業経

験がないのでできない。つまりデータ分析の力はあるのにビジネスを進められないみたいなのところがあります。そういう新卒の人たちに今、我々は何を最初にやっているかというところ、エンジニアリングスキルの向上なのですね。プロトタイプをぱぱっとつくってしまって、現場の人に持って行って、それでその反応を聞いてというところがやはり一番重要なことになるのではないかなと思っています。

というのも、たとえば、我々、不動産の業界なので、本当に今、橋本が言ったように、アナログな業界で、もうテクノロジーなんてまったく知らないよみたいな、体育会系の人たちが一番現場にいるわけなのですから、そういう人たちにAIでちょっと問題を解決するから何か教えてよみたいな感じで言っても、絶対に答えられないのですよね。なので、たとえば、「こういう分析をして、こういうツールをつくりたいと思うのですけれども、これは役に立ちますかね？」と言ったら、返ってくる答えは、「あれば使うよ」みたいな、そういうような答えしか返ってこないですね。なので、そういう聞き方は、ヒアリングの仕方でも駄目だし、どういうふうにやっていったらいいかというところ、最終的に、ぱぱっとプロトタイプをつくらせて、こういうものができますよという感じで見せてしまうということなのですね。そこで初めて、「ああ、こういう使われ方をしたら、これだったら結構使えるのではないかな」とみたいな建設的な意見が出てきて、それをキャッチして、それに対するフィードバックをしてみたいな形で、良いものをつくり上げていくという、たぶんそういう流れで。

里：そうですね、まさにその先ほど話した3年ぐらい前ですかね、AIがすごい流行って、我が社もAIをというかたちで上から下りてきて、AIというのを何か分かっていない人からの発注がすごく多くて（笑）。そのときに、おっしゃられている通りに、そのヒアリングから課題をつくったりとか、課題設定するというのは結構、筋が悪いですね。あるデータとか、ある程度のヒアリングから仮の問題設定をして、この問題でいくとこういうふうになるのですよ、という一回粗いものを見せた方がよい。そうするとAIで何ができるのかがイメージができて、ここをこうしたいんだよ、これがいいんじゃないか、というような意見がどんどん出てくるようになる。そうするとやはりちゃんと最終的にいいものができていくというか。

また、良いものを作り上げていくには、やはり試行錯誤が必要なんですけど、事業会社の方が、やりとりの回転が早い。クライアントとのやりとりだとやはりミーティングが早くてもウィークリなので、なので、それと比べるともう日々回転……

橋本：回転が早いと、あと距離が近い。ヒアリングももちろんするのですけれども、距離の近さを活かしてヒアリングの場以外でも相手を観察します。本当にヒアリングの通りなのかなと。ヒアリング相手が言うことを聞くのですけれども、一方でその言っていることの裏の本当の困りごととはなにかとか、実はこっちが本当の困りごとなのではないかなどと考えるのですが、その勘どころみたいなところがやはり必要だと思います。

福中：あと不動産業界のようなIT化の遅れたアナログな業界だと、これまで分析というものをあまりやってきていないことが多いです。そうすると、出てきた課題を解くのに大層なモデルはあまりいません。ディープラーニングももちろん使いますが、時間をかけてディープラーニングをやるよりも、可視化や回帰分析、決定木などの枯れた手法でクイックに対応することのほうが良い場合も多々あります。

なので、ある程度の分析スキルに加え、相手の困りごとを見つけるという出発点とそれを実装し現場に定着するまで並走、その2つがあればうまく回っていくような気がします。

石井：私の職場は社会人学生を受け入れる医療系大学院なのですが、入ってくる学生でスキルが高い学生がいて、最近では自分で勉強してかなりできるような人が入るのですが、応用とか、実際の研究展開とかになると、やはりそれなりに、運用能力とかが必要になってくるので、単にできるというだけではいけないということは感じますね。

次に加藤先生にデータサイエンティスト人材育成のお話を伺いたいと思います。

第3部：データサイエンス教育カリキュラムの動向

3.1 データサイエンス教育カリキュラムとは

石井：後半はデータサイエンスの教育カリキュラムに関して皆さんで議論していただければと思います。

加藤：正確には「大学のリテラシーレベルの数理・データサイエンス・AI教育」、と言うのですが、ここでは「データサイエンス教育」というふうに略させていただきます。今ちょうど、そのカリキュラム案が提示されていて、そのパブコメを募集しているところです（注：2020年3月25日に終了）。ですからもうたぶんこの号が出る頃には最終案は発表されていると思うのですが（4月15日発表済）。結構、急に出てきた話なのですが、なぜそれが、つくられるようになったかということを説明しますと、昨年（2019年）の6月に「AI戦略2019」が内閣府主導で発表されまして、その中に文理を問わず、すべての大学、高専生、これが大体、ここでは50万人と書いてありますが、統計を実際に調べると60万人に近い数になるのですが、その全員が初級レベルの数理・データサイエンス・AI教育を習得するという目標が明記されています。

しかもそれを大学、高専の卒業単位として認められる形で導入しなければならず、他方、優れたデータサイエンス教育プログラムを政府が認定する、つまりそれによってインセンティブをつけようという、そういう動きがあって、それでにわかに世の中が騒がしくなって、ありとあらゆる大学がデータサイエンス教育をやらなければいけないというふうな状況になってきているというのが現状なのです。

諸外国を見てみますと、やはりデータサイエンス教育というのはいろんな標準が発表されていて、主なところでは、昨年12月にACMがそのカリキュラムを出しました。それからヨーロッパではEDISONというところが出していますし、そんなかたちで海外では、我々日本に先行してカリキュラム標準を出しているという状況なので、日本も後れてはならじと今回発表したわけです。

ひと口にデータサイエンス教育と言ってもいろいろなレベルがありまして、今回発表しているのは、リテラシー教育ということで、これは全大学生、高専対象です。続いて今年度、策定する予定なのが応用基礎レベルで、これはダブルメジャーを想定してまして、たとえば、「経営学×AI」とか、あるいは「医学×AI」とか、自分の領域にAIを応用してデータに基づいた意思決定ができるような、そういう人間を育てようということです。これは、大体ターゲットとしては、

1年間、25万人と言っていますから、ちょうど半分ですね。それで単位数で言いますとリテラシーレベルは2単位から4単位ぐらい、ですから1科目から2科目ぐらいです。だからまあ大した話ではない。一方、応用基礎のほうは、もうちょっと多くて、6から8単位ぐらいになるかも知れないのですが、それぐらいの単位数ということになります。そして、さらにその上にはデータサイエンス学科とか、データサイエンス学部とか、それを専門で学ぶという課程もありますけれども、そっちの方のカリキュラム標準は今のところまだ手つかずというような状況ですね。

それが今の大学周りの状況なのですが、それに先立って小中学校の方でもデータサイエンス絡みの改革というのが始まっています。今年からは小学校でもプログラミングをやるということで、なかなか教えられる人材がいないので、どうやって教えるかということが大問題になっています。次に来年は中学校で、技術家庭の情報領域の内容が拡充されることになっていて、たとえば、ネットワークを利用した双方向性のあるコンテンツのプログラミングが導入されることになっています。今までよりも一歩踏み込んだ内容になっています。

続いて、再来年からは高校の情報Iという必修科目なのですが、その中にプログラミング、モデル型シミュレーション、ネットワーク、情報セキュリティ、データベース、情報デザイン、情報モラルといった、今、大学の情報リテラシー教育でやっているような内容の多くが高校レベルに下りてきます。ですから今、大学のリテラシーレベルのデータサイエンス教育のモデルカリキュラムが提案されていますけれども、これも長持ちしないですね。5年後にはこの、全員が情報Iを受けた世代が大学に入ってくるので、そのときにはこの内容は見直さなければいけないということになっています。

そのほかに高校では、選択科目になりますけれども、情報IIの中に、情報とデータサイエンスという内容が含まれてますし、また数学の方でも、必修科目の数学Iの中に統計的な要素、たとえば分散とか、標準偏差とか、サンプルとか、相関係数とか、そういったものが入ってきます。

というふうに、小学校から中学、高校、大学まで、データサイエンスの方向に、教育がシフトしてきているというのが現状です。

3.2 データサイエンス教育カリキュラムの概要

それで、今提案されている数理・データサイエンス・AIのリテラシーレベルのモデルカリキュラムの概要がどんなものかというのをお話しますと、4つのパートに分かれていまして、1つが「社会におけるデータの利活用」ということで、これは社会の中でAIがどんなふうに活用されているとか、AIを利活用するための技術だとか、それから、データ利活用の現場とか、最新動向とか、そういうお話ですね、社会の中で今データサイエンスやAIがどういう位置づけなのかというのを知るとというのがこの導入の部分です。

2つ目がデータリテラシーという部分で、ここは「データを読む」、「データを説明する」、「データを扱う」という内容で、表現はシンプルですが内容的には結構大きい部分ですね。統計の基礎に加えて、実データを、ここで想定しているツールはエクセルなのですが、実際に分析するとか、分析の方法だとか、可視化の方法だとか、グラフ化の方法だとか、そういうことを学んでいきます。ここはテクニカルな部分ということになると思います。

そして3つ目が、データAI活用における留意事項という部分で、ここは法律的な問題、倫理的な問題も含んで、個人情報保護とか、著作権だとか、あるいはAIの原則だとか、そういう、社会の中で、AIとか、データ分析を使っていく上で注意しなければいけないようなこと、それを学んでいくということが留意事項という部分になります。

それでこの3つのパートがコアな部分ということで、大体ここまで1科目2単位くらいの分量になります。ただし、大学のレベルとか、大学の学部によって、事情が違います。たとえば理系の学部ですと、データリテラシーのところは既存の科目でカバーできるということなら、分量がまた違って来るかもしれません。

それであとはオプションというかたちで、つまり選択ですね。これには、統計および数理基礎、アルゴリズム基礎、プログラミング、時系列データ解析、テキスト解析、データハンドリング、それから教師あり学習のデータ活用実践とか、教師なしのデータ活用実践とか、そういった項目が並んでいて、これは適宜、取捨選択して学習すればよいということになっています。

ということで、今大体の状況をお話しました。

石井：結局、データサイエンティストでは、今まで出てきている、情報科学とか、統計学とか、ビジネス応用とか、そういったものの学際的な学問という理解をしているのですけれども、実際のスキルとかは、あとの方で考慮されているということになりますか。つまり、具体的なスキルというのはオプションのところで対応するというふうな考え方に。

加藤：そうですね。ただ、正直申し上げて、このリテラシーレベルだけでは、なかなかスキルが身につくまでは到達が難しかろうと思います。やはり次の応用基礎ぐらいまでいかないとなかなか実際に使えるというところまではいかないのではないかなというふうには思っています。

吉野：実際に講義が始まるのはいつからですか。

加藤：2025年までの完成を目指していますが、認定制度は今年度からもう始まります。しかし、今年モデルカリキュラムが示されるので、1年弱の準備期間を経て来年度には多くの学校で実施されるのではないのでしょうか。

3.3 データサイエンス人材供給の状況と展望

橋本：ターゲット次第なのですが、このデータリテラシーが、応用までいかないと足りないというのはおっしゃる通りなのですけれども、たとえば企業でデータ活用を考えると、デジタル側だけではなくて、ビジネス側もできないと駄目なので、こういうビジネス側の人たちにとっては、このデータリテラシーというのはきっと役立つと思います。それこそ放送大学の身近な統計、100回ぐらい見せたいのですけれども（笑）。1回見ればいいのですけれども、なんかそういうのにはこれはきっと役立つのではないかなと。

デジタルの部署だけが頑張っても駄目で、リアルなビジネス現場の部署も頑張ってもらわなければ駄目なのですけれども、そういうときにはこのデータリテラシーみたいなのはちょうどいいのではないかと。

ビジネス現場の方に高度な機械学習のアルゴリズムの説明をするよりも、グラフの見方とか、平均だけでは駄目で、分散や中央値も大事だよとか、そういうほうがインパクトはでかいのではないかなと感じます。

グラフの読み方すら危ないところが、下手すれば、デジタル側のグラフ書かせるほうもね、都合のいいグラフを描いてしまったりして、だからギャップはあります。

加藤：意図的に都合の良いグラフを描けるとよいのですけれどもね（笑）。

橋本：頭が良い人は逆にそういうことをするかもしれません（笑）。

石井：実際のこの教育を受けた学生が社会に供給されてくるのは5年先とか。

加藤：早ければそういうことになりますね。ただし、認定制度はもう、20年度の夏頃から公募が行われる予定なので、先行して始めている大学もあるということですね。

橋本：滋賀大の一期生が就職活動していて、何人かお会いしました。

加藤：内容は大学ごとに若干違うのですけれども、もうすでに全員がデータサイエンス教育を受けるということを実施している大学はいくつかありますので。

橋本：滋賀大、横浜市立大、武蔵野大などはデータサイエンス学部ありますね。あとは筑波大、これから全学生にやりますよね。

加藤：北大もそうですね。

石井：もうすでに人材供給が徐々に始まっているということでしょうか。

加藤：たぶん（笑）。どの程度うまくいっているかというのは、評価は待たなければいけないのですが。

里：実際に来ている方はどんな感じなのですか。

橋本：難しいのですけれども、本当に二極化という感じで。

石井：現実に入ってくる新卒というのはもうかなり高いスキルを持っているというか、そういうケースもあり得るのですか。

橋本：スキルの言葉の示すもの次第ですが、ビジネス課題解決だとあまりないですけれども、機械学習のアルゴリズムなど詳しい方は結構いらっしゃいます。

里：元々専門とか、自然言語処理とか、そういう研究をしている方とかだと、もう技術を持っていたりとかするので、そんなにまだ変わっていない感じですかね（笑）。

第4部：総括

石井：それでは、ゲストインタビューの方々の発言が一巡しましたので、総括として吉野さんから何かご質問ございますか。

吉野：データサイエンティストのスキルだけではなかなかビジネスはまわらないというお話がありました。システム化するといったときに、いろんなケースがあると思うのですが、既存の本業の情報システムにデータサイエンスなり、AIの部分を追加していくという感じが多いのか、それとも既存の SoR（Systems of Record）はそれはそれで置いておいて、SoE（Systems of Engagement）の新しいシステムを開発するという感じが多いのでしょうか。

里：分らないです。両方あるとしか言えないですね。

加藤：結構前からIT技術者が足りないと言われてきていて、IT技術者の中でも特に足りないのは上流工程をやる人だとかというような話と平行な話のように私には聞こえたのですが、

橋本：言葉が、DXなのか、データサイエンスなのか、AIなのか、IoTなのか、主語が変わっても、なんか根っこは同じような気がします。

吉野：上流の、そもそも何のためにシステム化しているのかといった課題は、システムの問題ではなくて、その企業の経営そのものの課題ということでしょうか。

橋本：そうだと思います。

吉野：何のために会社をやっているのかという、そこが明確になっているのかという、そういう問題になってきてしまうような気がします。

あと、なるほどなと思ったのは、事業会社で、自分のデータを使って、それで自分の事業をどう良くしていくかという方がデータサイエンティストが活躍しやすいというお話です。里さんは両方の立場を経験されていて、自分でまわしていたこともあるし、外部の立場で分析するということが現在はやっておられる。

人材育成の観点でその2つの立場の違いについてもう少し伺えますか。

里：データサイエンティストでしっかり実力をつけるという意味では、一回、事業会社みたいなところで腰を据えてまわしていくという経験はあったほうがいいとは思っています。どんどん改善していくというのを積み上げていって、成功体験も、失敗も含めて、やっていくというのはすごい貴重な体験なので、やった方がいいなと思っています。

一方で、人によると思うのですが、同じデータだと飽きてしまう（笑）。それで、コンサルティング会社のいいところは、いろんなデータ、ビジネスを知ることができて、そこでチャレンジできるというところ。視野が広がるのでこれもあったほうがいいと思う。なので、僕としては両方やった上で自分に合っているものでいったほうがいいのではないかなとは思っています。

福中：どっちが先なのですかね。僕自身もデータ解析コンサルをやっているんな分析といろんなデータを知ってから今の事業改善で仕事をやっている感じなのですが、逆もありのような気もしていて。

橋本：前後ないと思いますよ。深く縦にか幅広く横にかの話ですので。結局、面積、広く取れば、大きく伸びるという。

福中：関与度は深いけれども、幅が狭いか、幅は広いけれども、関与度が浅いか。

石井：あとデータサイエンスの技術的な、あるいは、学問的なものは完全に出切って、もうピークに達していて、もうこれ以上出ないような感じですか、それともこれからまたいろんな展開が出る。

里：手法とか、アルゴリズムの話ですか。

石井：そうです。

里：どんなものかは分かりませんが、出てくるんじゃないかなと思います。今から新しいデータがどんどん出てきますし、それに伴って技術、少なくともデータをさばく方はどんどん発展していくと思いますね。今のデータ量はさばけるけれども、次のデータ、次の時代のデータはさばけなくなると思うので。その上でそれだけ増えたときに、今までの既存の手法で全部解決できるのかと言ったら、きっとそうではない気がするので、また別のものは出てくるだろうなと思います。

石井：時間になりました。今日の座談会インタビューはこれで終わりたいと思います。まとめですが、本日はソフトウェアジャパンのプレゼンタ、デジタルプラクティスの執筆者、関連委員会などで、ビッグデータ、AI、IoTに関係してこられた方々に、現在のデータサイエンスを巡るトレンドと今後の展開について、また、人材育成について、お話をいただきました。今後、データサイエンス、ビッグデータを巡る状況は、どんどん展開していき、発展していくことは間違いないと思います。その中で、皆さん方のご健勝とご活躍を祈りつつ本座談会インタビューを締めたいと思っています。今日はどうもありがとうございました。



インタビューの様子：右から時計回りに：吉野氏、加藤氏、福中氏、橋本氏、里氏、石井氏
2020年3月17日13:00～14:00情報処理学会事務局会議室にて