

座談会

「ビッグデータのデータサイエンス ～ニューノーマル時代のビッグデータ～」座談会

進行役：里 洋平（（株）Village AI／nat（株）／（株）Lupinus）

インタビュイー：高柳慎一（（株）ユーザベース），安部晃生（（株）コネクトデータ），飯尾 淳（中央大学），牧山幸史（（株）ヤフー）

インタビュアー：石井一夫（公立諏訪東京理科大学）

本特集は、「ビッグデータのデータサイエンス」というタイトルで、ビッグデータを対象としたデータサイエンスについて、特に、コロナ禍や気候変動時代におけるビッグデータのデータサイエンスの在り方を意識しながら企画した。それを受けて、今回の座談会では、本会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（略称：ビッグデータ研究グループ）の運営委員メンバにより、「ニューノーマルにおけるデータサイエンス」と題して、最新の関連トピックについてお話しいただいた。本企画が、日々、目まぐるしく社会状況が変化していく中での、データサイエンスの今、これから、について、日々の業務のヒントになれば幸いである。



里 洋平（正会員）（（株）Village AI／nat（株）／（株）Lupinus）

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、DATUMSTUDIO（株）を設立。2021年7月に退任し現在は、（株）Village AI 代表取締役、nat（株）取締役、（株）Lupinus 社外取締役。本会ビッグデータ解析のビジネス実務利活用研究グループ幹事を兼任。



高柳 慎一（正会員）（（株）ユーザベース）

2020年総合研究大学大学院複合科学研究科統計科学専攻博士課程修了。博士（統計科学）。2020年（株）FORCAS入社。2021年統合により（株）ユーザベースへ転籍。B2B事業向け顧客戦略プラットフォームFORCASの開発に従事。徳島大学客員准教授。本会ビッグデータ解析のビジネス実務利活用研究グループ幹事を兼任。



安部 晃生（非会員）（（株）コネクトデータ）

（株）コネクトデータ代表取締役。企業におけるデータ利活用のためのコンサルティング、分析、開発、教育に従事。



飯尾 淳（正会員）（中央大学）

中央大学国際情報学部教授。人間と情報システムのインタラクションに関する研究に従事。特定非営利活動法人人間中心設計推進機構理事。

（一社）ことばのまなび工房理事。博士（工学）、技術士（情報工学部門）、人間中心設計専門家。

牧山 幸史（非会員）（ヤフー（株））

ヤフー（株）にてデータサイエンス業務に従事するかたわら、（株）ホクソエム代表取締役社長と徳島大学客員准教授を兼任する。



石井一夫（正会員）（公立諏訪東京理科大学）

公立諏訪東京理科大学工学部情報応用工学科教授，久留米大学医学部内科学講座心臓・血管内科講座客員准教授，少子高齢化および地球温暖化問題の克服に向けた医療ビッグデータ，環境・農業ビッグデータの教育研究に従事，本会ビッグデータ解析のビジネス実務活用研究グループ主査。

里：こんにちは，本企画でゲストエディタをやらせていただいている里です．本日は，よろしくお願いたします．最初に，自己紹介を皆さんにお願いしたいと思います．高柳さんからお願いしてもよろしいでしょうか．

高柳：（株）ユーザベースでデータサイエンティストをしている高柳です．よろしくお願いします．

今メインでやっているデータサイエンスの業務は，営業支援システム開発です．本特集の論文にも書かせていただいたのですが，企業が営業するときにはアタックリストを機械的にAIで作るようなシステムを作っています．よろしくお願いします．

里：よろしくお願いします．

次は，安部さん，お願いいたします．

安部：（株）コネクトデータ，代表取締役の安部晃生です．

私は，普段はクライアント企業のデータ利活用を支援するためにコンサルティング，分析，開発，教育等々をやっております．最近だとオープンデータ活用みたいところに非常に興味がありまして，世の中のオープンデータの流通を活発にしたり，それをベースに何か技術発展したりのようなことを狙って，delikaというオープンデータのプラットフォームを開発，提供をしています．よろしくお願いします．

里：よろしくお願いします．

次，飯尾先生，お願いいたします．

飯尾：飯尾でございます．どうぞよろしくお願いいたします．

私は中央大学の国際情報学部というところで教鞭を取っていまして，その国際情報学部というのはほかにあまりない学部名なのですけども，情報系と，あと法律の先生たちがいます．今の情報社会を技術的に支えるのが情報系の学者の分担であって，それを社会実装していくときに，社会のルールに合わないといけないよねということで法律系の先生方が，自動運転の車が事故を

起こしたらどうなるのみたいな、そういう話は今後、重要になってくるので、そのようなことを学生に教えているという、そういう建て付けの学部なのですね。私は法律ではなくて、情報系のほうで、いろいろ教えています。

データサイエンス関連でいうと、今日の座談会の中でお話できると思うのですがけれども、本学は今すごくデータサイエンス教育に力を入れているというか、政府の、ちゃんとやれよというのに乗かってやっていますので、そのあたりのお話なんかをできればいいかなと思っています。

里：ありがとうございます。

牧山さん、お願いします。

牧山：ヤフー（株）の牧山と申します。仕事ではデータ分析全般をやっていて、何でも屋みたいな感じです。よろしくお願いします。

里：はい。よろしくお願いいたします。

ありがとうございます。

では、最後に石井先生、お願いします。

石井：公立諏訪東京理科大学の石井と申します。今年の4月から現職に異動してきました。それ以前は久留米大学にいて、そこで医療ビッグデータを中心に教育研究をやっていたのですが、現在は本学工学部の情報応用工学科に勤めています。本学は全学を上げて、AIとか、機械学習とか、ビッグデータとかに力を入れていて、私もそれに乗っかっていろいろやるということで、相変わらず医療ビッグデータを中心に教育研究をやっています。

最近は地球温暖化にも興味があって、近ごろすごい豪雨とか、熱波とか、大変なことになっていますけれども、そういう関係の分析も含めいろいろやっています。

今日は本企画の、コーディネータとして本座談会をバックアップさせていただきたいと思います。よろしくお願いします。

里：お願いします。

データと法律

里：どなたか、これをぜひ話したいという強い思いがあれば、その話からやっていければと思うのですが。

安部：さっき飯尾先生から法律と情報みたいな話があって、私も結構、興味がある分野で、先ほどの当社のサービスでdelikaを作ったきっかけの1つとして、著作権改正みたいなところがあるのです。日本の著作権法というのはいわゆる情報処理の用途で非常に自由に利用できるような形に改正されて、日本は機械学習天国だ、みたいなことがいわれることもあると思っています。

そういうデータの使いやすさというものが法として整備されている一方で、そのデータの流通みたいなのところというと、こういうデータを勝手に使っていいのだろうか、そういう自制心みたいなものが働いて、データの活用みたいなのが活発になりづらいのと、サービスを運営しているところがあります。皆さんそのデータの活用というところで、恐らく自分たちが獲得しているデータに関しては問題なく使っていると思うのですが、オープンデータにしる、いわゆる売買されているデータにしる、外部のデータに関して、どういうふうな考え方をお持ちなのかなというのに、少し興味があって、お聞きしたいなと思います。

飯尾：では、それを受けてちょっとしたエピソードというか、学生に伝えたいことという観点から、グレーゾーンというわけでもないのですが、外部からデータを取ってくるときに微妙なところを、超えてはいけないところはどこにあるのかというようなところが少し曖昧ではっきりしていないケースというのは結構ありますよね。そこがやはり気になりますね。

有名なところだと、昔、Librahack事件というのがありましたよね。岡崎市の図書館に1秒ごとに、あれは確か1秒ごとにアクセスしていたと思いますけれども、図書館のシステムのほうが、作りが不十分でなんかどんどんリクエストが溜まって行って、それでダウンしてしまったみたいな。1秒ごとぐらいのアクセスだったら全然オーケーなのではないかと普通は思いますけれども、図書館側のシステムがそれに耐えられないようなシステムだった。そういうことは、普通は技術屋としては想像できないですが、実際に、ああいうことが起こってしまうと、それは技術面からの問題もあるけれども、法律も整備しきれていないところですし、社会の文化的なところも、あれは大変良い教訓を残したと思うのです。そのあたりを学生にどう伝えていくのかというのは少し気になっていますね。

今、私も、Twitterのデータを取ってきて、それで毎日分析というか、20分おきにデータを取ってきて、それで分析しているシステムを運用して、いろいろやっています。TwitterのデータはAPIを叩いて取得していますが、それ以外のデータの収集がなかなか難しい。そのあたりもデータを取り扱う技術というよりも、社会的にどこまでやってよくて、そこから先はアウトという話で、我々は経験上、ここまでだったら大丈夫だろうなという感覚は持っていますが、そのあたりを、学生にどう教えていくのかというところはかなり気になっています。

里：ほかの方はどうでしょうか。

石井：気になっているのは、私は医療ビッグデータを使っている関係で、この分野は個人情報保護の規制がかなり厳しくて、使うデータが個人を特定できないようにするとか、データ分析というのは個人情報との戦いみたいなところがあり、どこまで個人情報を暴くかというところで、個人を特定できるぎりぎりまで攻め込んでいくというのは結構やるのですが、そのあたりが法律と、あと倫理的なところとの綱引きというのが結構大変だなというのはありますね。

飯尾：今、倫理という言葉が石井先生からありましたけれども、最近やはり、研究倫理の問題がすごく、ややこしくなっています。バイオの研究とか、石井先生は割とそちらのほうもずっとやられてきたので医療の土地勘はあるかと思いますが、人造人間を作ってはいけないとかね（笑）、試験管ベビーがどうだとか、そういうような話は、これはあかんやろという感じ

で、すぐ分かるのです。けれども、だんだん最近では周辺領域というか、つまり社会科学のところまでそういうのが求められるようになってきて、昨今、とうとう我々も白旗を上げて、その波に飲まれているかなというところがあります。

私の周辺でいうと、文学部の心理学の先生たちがだいぶ戦ってくれたのですけれども、世間の波に飲まれてしまって、少しやりすぎなのではないかなというふうには思っています。要するに、人にかかわる研究というのはどこまでの倫理を求めるんだと。個人情報扱ってれば、それは人にかかわる研究だってされてしまうのですよね、今の文脈だと。そうすると、研究倫理委員会を通さないといけないとか、面倒くさいことばかり増えて、それはしょうがないのかなとは思いつつ、困っていますね。

石井：新型コロナウイルスのデータ分析とか、まさにそれですね、新型コロナウイルス感染症患者のデータ分析とか。

飯尾：おっしゃる通りだと思います。だから、誰かが反旗を翻してくれないかなと思っているのですけれども（笑）、残念ながら私はそこに、先頭に立つ勇気がないので（笑）。まあ、そんなところですかね。

データの流通

里：ほかの方、どうですか。

高柳：だいぶ視点が違っていて最近よく考えている会社でも話している内容ですけれども、外部データとか、オープンデータに依存してしまうと、それは経営リスクだよね、みたいな話はよくしていますね。たとえば、データ分析というのは要するにデータの加工産業みたいなものじゃないですか。データを仕入れてきて、それを適切に調理して、料理として出すのが分析レポートですし、それをシステム化したりしているわけです。こういう状況で、たとえば今年だと特に野菜が値上がっているかと思うんですが、同様に外部データとして買っているものが値上がってしまったとか、データそのものの供給を止めたとかとなると、我々のそのデータ加工ビジネスが、ぼしゃってしまうのでそれは経営リスクだよね、じゃあ、内製化するか、どうしようかな、みたいな視点で話していることが結構多いですね。

オープンデータの値上げというよりも、どちらかという、供給がいきなり止まって、代替先を探さなければというので、慌てふためくのがちょっと嫌かもみたいな話はよくしていますね。

安部：気になる場所としては、オープンデータに近い概念でオープンソースというのがあってと思うのですが、世の中のシステムとかというのは結構オープンソースのコードとかを使って、まわっていたりするではないですか。オープンソースとオープンデータの性質の違いというのはどういったところにありますか。1つの違いは、継続的に更新されていくみたいなことだと思いますけれども。

高柳：我々のビジネスの話をしてしまうと完全にそれで、オープンソースだと適当にバージョンというか、GitHub、Gitだと、コミットIDとかタグで固めてしまったものさえあれば、しばらくはまわるのですが、オープンデータはイメージとして常に新鮮なデータが入ってきていないと

ヤバいみたいな話なのですね。たとえば、今、まさに我々、Zoomで会議していますけれども、Zoomを使っている企業というのをデータの中から抽出したいみたいな案件があって、それは毎月毎月、毎時毎時、新しくZoomを使うという企業があったり、やめてしまったりみたいなものがあるので、そういう意味で、継続的に見続けてなければいけないので、ちょっと違うみたいな、ストックというよりもフローに近い感じなのですよ。そこに差がある感じですね。

安部：面白いですね。確かにおっしゃる通りかなというふうに思います。それは結局オープンデータというものの外部依存というところが、あるいは、外部から一定に供給されることが止まるリスクというものが上手くコントロールできれば、まわるという形になる。

高柳：そうです。おっしゃる通りです。

安部：それは各々のビジネス主体がもう独立に動いているから、その供給するという関係性を意識せずにまわすので止まるリスクになるという感じですかね。結構、世の中というのは、どこかが止まるとほかに影響が出るから、それを止めることに対してのリスクを、公共機関とかが守ってくれたりするではないですか。あるいはビジネスサイドでも、うちのこのビジネスは赤字だからやめたいけれども、これをやめると影響が出るから困るよね、みたいなところがあると思うのですよ。

でも、データだと結構それがリスクになり得るということは、あまりデータの流通というもののというのが世の中に意識されていないのかなというふうに思いましたね。

高柳：そうですね、流通が意識されていない、ありそうですね。

データの質

飯尾：オープンデータに関してはね、質の問題も結構あるのではないかなと思うのです。そのあたりは皆さん、いかがですか。つまり使い勝手のいいというか、いわゆるTimothy "Tim" John Berners-Leeの5 Starオープンデータでいえば、レベル1とか、レベル2のところでとどまっている。少なくとも私の経験では、使い勝手の悪いデータばかり流通していて、そこをなんとか加工して、クリーニングして使っているというような状況なのですけども、皆さん、いかがでしょうか。

安部：私もそう思いますね。あの5 Starでいうと、機械判読が可能みたいなレベルが確かレベル2か、3ぐらいにあったと思うのですが、まずはそのレベルに到達するところが最初かなというのは思っています。今、当社が提供しているdelikaというプラットフォームも実は、5 StarだとRDFを使って、自由にデータとデータが繋がるよ、みたいなところがあると思うのですが、それよりも実際に利用できる部分をまず目指そうよみたいなところで、機械判読のデータというものを流通させたいなという思いがありますね。

データが使いにくくて、データサイエンティストがみんな同じデータの前処理をしているみたいなところがあるので、まずそういうところをなくすことによってデータの価値というものが社会的に認知されるようになってみたいところを目指していきたいなと思っています。

飯尾：世界中のデータサイエンティストが圧倒的に前処理に時間を費やしているというのはものすごい時間の無駄というか、生産性を下げていますよね。そこは改善したいですね。

安部：カレンダーのデータとか、たとえば、Googleカレンダーから取ってくるだとか、気象庁のあの汚いデータを引っ張ってくるとか、表形式になっていないCSVとして扱いづらい政府データを使って、みんなされていると思います。日本の生産性を国が積極的に落としにいつているという、ひどい状態になっています。

飯尾：いや、みんな同じ思いなのだなと思って、共感しました（笑）。

里：牧山さんは何かありますか。

牧山：ヤフーはいろいろなサービスを展開しているのですが、社内のデータに関してはかなり整備されていて、どのサービスのデータがどこにあって誰にアクセス権を申請すればいいかなどが一覧で分かるようになっています。

しかし、たとえば、同じグループ会社のPayPayとかのデータに関しては、複雑な手続きを経ないとアクセスできないという問題があって、かなり苦労してデータを手に入れないといけないので、そこが障壁になっています。

それで、私が最近注目しているのは、Federated Learning（連合学習）と言って、それぞれの組織はプライベートなデータを公開せずに機械学習モデルを作るという手法なんですけど、それに注目しています。たとえば、WeBankという中国のデジタル銀行が多重債務者を判定するのに機械学習を使っていて、それはFederated Learningを使って、ほかの銀行のデータと照らし合わせて多重債務者を判定するのですが、そのほかの銀行のデータというのはそのWeBankがもらっているわけではないのです。共同で多重債務者を判定する機械学習モデルを作っているという感じです。そういう仕組みに今ちょっと興味を持っています。

飯尾：それというのは何か業界団体みたいなものがあるって、そこに加盟している各社が自分たちのデータは全部には公開しないけれども、共通するモデルを、自分たちが持っているデータで、それぞれが上手いこと協調させて、学習させて、1つのモデルを作るとか、そんなようなイメージなのですか。

牧山：業界団体があるかどうかはちょっと分からないのですが、イメージとしては、自分のデータだけで学習したときのモデルのウェイトだけを共有しましょうと。そしてグローバルなコンセンサスを持ったモデルを作りましょうというようなイメージで考えていただけると分かりやすいかなと。

高柳：今、Zoomのほうにリンク^{☆1}を貼ったのですが、まあ、Googleを筆頭に、おっしゃったような個人情報のターゲティングをもうやめようよという技術として、牧山さんが言っていたFederated Learningは、今、送ったリンクだと、FLOCとかと略されてしまっているのですが、これ、Federated Learningの頭がFLですね、という技術が台頭してきている感じは、確かに印象は受けます。

飯尾：このFLOCというやつ、あまり評判がよくないみたいですよ（笑）。私も、よく知らないのですけれども、学生がFLOCは今後スタンダードになるんですかとか聞いてきて、ちょっとだけ調べたことがあって、なんか評判悪いらしいよというような話をした覚えはあります（笑）。

安部：そのFLOCの問題の1つとして、学習データの偏りみたいところで、Federatedする相手先のその属性によってモデルというのが改善されていくので、偏った集団に対して学習してしまうと、公平ではないAIができあがるみたいなことは1つ問題として挙げられているかなと思っていますね。

企業の持っている集団で学習するという考え方でいうと、ビジネス上はまわりそうな気がしますが、たたとえばGoogleみたいな大きな企業になると、その偏り自体が問題視されるという話があるかもしれません。

牧山：公平性の問題は普通の機械学習でもあるような気がします。

飯尾：それこそ、技術的な話題というよりは、参加している企業の間での調整をどうするかみたいな、社会のルールがまだ未整備的な、そんな話題というふうに捉えられますよね、この問題は。

牧山：そうですね。それはあると思います。ちゃんとしたガイドラインを作らないとなかなか実現するのは、難しいかなと思っています。

人材育成

里：では、ちょっと話題を変えていきたいのですが、冒頭であったのは、人材育成だったり、あとニューノーマルなデータサイエンスだったりとか、そのあたりのお話があればなと思っているのですが、人材育成で最近の状況や、思っていること、考えていることなど、何かあれば、皆さん、お願いいたします。

飯尾：若干ちょっと宣伝めいて恐縮なのですが、文部科学省が主導しているのでしたっけ、今、全国の大学でAIとか、データサイエンス教育をばんばんやれみたいな、そういうプログラムが動いていまして、そこに私どもも参加しています。認定^{☆2}を受けるのを目標にしてどうのこうのなんていう話をしているのですけれども、全学でAIデータサイエンス教育をしようと。

それは、なかなか挑戦的なことで、中央大学というのは割と文系寄りの大学で、全学対象でAIだとか、データサイエンスのリテラシー教育をやると、そういうようなことをやっています。

それで、全学対象で、うち、8学部あるのですけれども、私どもは学際ということで理系と文系が融合したところで、理工学部以外はほぼ文系なのですが、法学部とか、商学部、経済ですね。なので、学生の8割方であるそういう学生に向けてAIとか、データサイエンスの教育をするんだと。もちろん、通りいっぺんのお話で終わってしまうかもしれない。ほとんどの学生はね、

今の社会はAIに支えられているのだよみたいな、そんなようなお話で終わってしまうのですけれども、そんな中で私どもが、それと理工学部にも情報工学科がありますので、そんなようなところで少し突っ込んだ教育をするのかなという、そういう建て付けでやっていますね。

これが、困ってしまうのは、AIデータサイエンスリテラシーレベルというのが、今、国からの、なんていうのですかね、プログラムの指針として出されて、さらにその上に、応用基礎の認定プログラムというのが、今、検討中らしいのですけれども、知り合いの先生もその委員会に入って、がちゃがちゃやっているというふうに聞きました。

石井：応用基礎のカリキュラムそのものは、2021年の3月に公開されています^{☆3}。

2～3月にパブコメ^{☆4}があって、その直後にリリースされたと思います。

飯尾：応用基礎の認定プログラムというのが細かいところがまだ決まっていないので、どうのこうのなんていう話を1カ月前か、そのぐらいにやっていました。

それで、それが全学の50%以上が履修しないといけないとか縛りがあるらしくて、さすがにそれは難しいだろうと（笑）。うち、3万人からいますからね、そのうちの8割ぐらいが文系のはずなので、8割、9割ぐらい？ まあ、8割かそこらですよ。なので、さすがにちょっとそれを目指すのは厳しいだろうと、私は個人的には思っているのですが、志を高く持てなのか知りませんが、そんなようなところを最終的なゴールとして、このAIデータサイエンス全学プログラムにかかわっているというか、旗振り役の先生は頑張っています、というご紹介でした。

これらについて質問があれば、できるだけお答えします。いかがでしょうか。

石井：企業とのかかわりとか、そういうのはありますか。

飯尾：ありがとうございます。

全学教育なのですがけれども、バックエンドにAI・データサイエンスセンターという部署がありまして、私もメンバの1人なのですが、そこは企業とタイアップしていろいろ共同研究をやりましょうという話は、別途進んでいます。もちろんそこで得られた知見なんかも、その全学プログラムのほうにフィードバックしていくとか。

あとは、何だろうな、AIデータサイエンス総合という科目ですね、企業からの先生をお迎えして、それで最先端の話をしてもらってコマとかもあったと思いますね。やはり大学だけだと難しいですよ、企業と連携してやらないと。先ほどのオープンデータの話もありましたけれども、やはり大学だけだとリアルなデータというのはなかなか持っていないので、企業さんも、出せるデータと出せないデータ、当然、あると思いますけれども、出せる範囲でリアルなデータとか、事例とかを出していただくと、学生もいい刺激を受けますので、そんなようなことも入れていますね、プログラムの中に。

里：実際、企業の中での人材育成みたいな話を、高柳さんとか、何かありますでしょうか。

高柳：今だとすべてがリモート前提になってしまって、いろいろすごい教えにくいなというのはたぶん皆さん同意なところだと思うのですが、まずそこが1点あります。企業自体ではなく、企業と大学、また学生さんとかかわりあいのお話ですと、私もちょうど学生さんに講義をさせてもらう機会があるのですが、学生に何を教えるといったときに、意外と学生さんは皆さん野望がないというか、何だろう、学生感がなくて、みなさんととても達観されているというか大人で（笑）我々がデータサイエンスをやるとこんなに楽しいですよみたいな話をしても、なんか別世界のような話に感じてしまうのがなんか問題だよという話をしていて、じゃあ一体何を教えたらいいんだろうみたいな（笑）、データサイエンスがあたかも自分とはまったく関係のないマンガかアニメの世界の話を聞いているような印象になってしまっているっぽいのでそこをなんとかしたいなど、そのギャップをどう埋めようかなというのが、学生さんと企業の間でかかわっていかうと思ったときの課題ですね。

実務面での人材育成はやはりリモートが多くて、エンジニアリングの話だとOSとか、オープンソースとかの話もあったように、GitHubを使って開発するやり方をすれば、大体みんな非同期に開発できているから、まあ、そこはいいのだけれども、実際にデータを扱ってうんぬんとか、細かい個別の相談まわりの話になると急にやりにくさのギャップが出てきて、いろいろZoom的なツールを使って、いつでも気軽に声をかけられる状態にしつつ進めていますみたいな状態にもしているのですが、やはりまだ対面でやっていたときに比べるとギャップがあって、いかなものかなと思っているというのが正直な感想ですね。

里：ほかの方、どうですかね。

牧山：AI人材と言ったときに、AIを研究できる人なのか、AIを実際にサービスに利用して運営させることができる人なのか、ちょっと定義が曖昧だなと思っていて、企業で必要になっている、需要が高いのは、研究する人よりは、プロジェクトを率いてサービスに機械学習を入れ込む人たちです。そこら辺がすごく今のところ人材が少ないので、ぜひ大学とかで教育していただけるのであれば非常に助かるなという感じですね。

石井：前処理をする人という意味ですか。

牧山：前処理も含むという感じですかね。実際の機械学習のプロジェクトというのは、割と工数が読めなかったり、どれぐらい成果が上がるのかというのが分からなかったりする。そこら辺を上手く進める、ちゃんと技術選定とかもして、どういうモデルを作るのか、どういう指標を見て成功を判断するか、サービスに入れるかどうか、入れることによってコストがペイできるのかとかを判断して、それで導入するかどうかを決めるとかも必要だし、いろいろな能力が必要になってきていて、そこら辺の人たちがいるとすごく助かるなという印象です。

飯尾：おっしゃることはよく分かるというか、企業さんはそういう人材を欲しがらるだろうというのは、私も昔、企業におりましたので（笑）、分かるのですけれども、結構、今、求められた能力というのは、経験によるのではないかなという気がします。大学で、短い4年間しかない大学生活の中で、しかも1年生なんていうのはもう高校から上がってきて、本当に右も左も分からないと言ったら怒られますけれどもそういうようなので、実際に、AIとか、深い学習ができるというと、本当に3年、4年生になってからだと思います。そうすると2年間でどこまで経験を積めるかということ、なかなか難しいですよ。

難しいであろうということは分かっていつつ、先ほどの全学のあの教育の枠組みの中で、別途、私が今、持っているゼミとは別に、全学対象のそういうデータサイエンス系のゼミを来年から担当するのですが、そこで実際のリアルなデータを扱って何とか少しでもそういう経験を積ませるような教育ができればいいかなという挑戦はしようとは思っています。ただ、どこまでできるかなというとなかなか厳しい（笑）。厳しいリクエストです（笑）。

牧山：なるほど。少しでもAIプロジェクト推進の経験があれば、さらにそのAIの仕組みとかについて知識があれば、十分かなと思っています。

里：安部さんは外から人材育成を支援するということはあったりするのですか。

安部：今まさにとある企業に半年ほど毎週1時間ぐらい講義形式でやっていたりするのですが、それをやっていて思うところは、企業にいる方というのはいろんな立場の方がいて、当然、文系の方もいれば、理系の方もいらっしゃるというので、やはり知識レベルがばらばらかなというところがありますね。そういう意味では、先ほど飯尾先生もおっしゃっていたみたいな、勉強教養レベルでミニマムラインみたいなものを定義していただいて、それでそれを修めたぐらいの人材としての前提で話せるとまた話せることが違うことが多いのかなと思っています。

それこそ、さっき牧山さんがおっしゃっていた、AIを運用するとき、プロジェクトのまわし方、精度がどのくらいになるかといった読みづらいところもあるし、経験でしか積めないところではあります。なので、企業でしかできないことは企業にやらせるという前提で、大学における位置づけとしては、大学ならではの教養、基本的なところでいうと線形代数だとか、微積だとか、そういったところのミニマムの感覚みたいなものを育ていただき企業に来ていただけると非常にやりやすいかなと思いますね。

まさにデータサイエンスというのは、サイエンスの領域に限らず、ビジネス領域も含めて、広範の領域の総合格闘技なので、あらゆるスキルを身に付けている人というのはいないと思いますが、とは言え、最低限のコミュニケーションできるレベルのというのはあるかなと思います。データサイエンティスト協会のスキルチェックリスト^{☆5}をベースにされていて、その中に★1, 2, 3というのがあるのですが、その★1をベースに教育するだけでも、結構、苦労しているので、企業に来る人間が当たり前のレベルになっていると、企業の中でそのデータ活用人材というのがどういうふうに活躍していくかというのは、まさにビジネスに特化した形で進められるのでいいかなというふうに思いますね。

里：ありがとうございます。この話題に関して、何かほかにご意見などあれば、お願いいたします。

飯尾：今おっしゃった中で、線形代数は重要だね、みたいな話は、まったくその通りで、今、プログラムで、ライブラリとかを使えば、簡単にAIみたいなものというのは作れるようになっているので、では特徴量は何にするのみたいな話をしたときに、これは笑い話で、もう卒業してしまったうちの学生ですが、なんか3つぐらいパラメータを入れていて、どうしても性能が上がらないんですよとか言っていて、何を特徴量に入れているのと言ったら、1つ目は、これこれで、もう1つ目は、これこれと、似たような特徴量を入れているので、それ、いいのかなと思います。聞いていたら、それで、3つ目は何かと言ったら、その平均値を入れています。

と、おいおいおい、そんなね（笑）、従属変数を入れてどうするよみたいな、笑い話がありましたけれども、そういうところをしっかりと大学としては教えていかないといけないなどは常々思っています。

安部：最近、ディープラーニングをちょうど教える機会があって、その線形代数の知識が課程になっていれば、活性化関数みたいなところで非線形関数を食わせないと全部で見たときにもう線形結合の結合だから、線形結合になってしまうよねみたいなことを説明するだけでも、結構、苦労するので、そのくらいの常識感があるといいかなという感じがします。

用語として、エンジニアはエンジニアが使う用語があるし、サイエンスの人はサイエンスで使う用語があるし、ビジネスの人はビジネスが使う用語があります。そのあたりのベースラインみたいなものが業界の中で統率が取れるとたぶんお互いにコミュニケーションが楽になります。得意分野は得意な人がやればいいのですけれども、それをコミュニケーションするための最低限の教養があるといいかなというふうには思っています。

里：高柳さんのところとかは、研修というのはどんなふうに行われているのですか。

高柳：OJT（On the Job Training）一択という感じでしょうか。そもそも新卒採用を積極的にやっているわけではなくて、ほぼ中途採用だけなので、そういう課題はないです。逆に、私が副業でやっている会社でデータサイエンスに関する研修を提供するとなると、どちらかというと、さっき安部さんが言っていたような、★1個の内容を教える、みなさん大学生のときに意外と学んでいないのだけれども、今になって、やらなければならぬみたいな話とか、統計学や学問のありがた味が今になって分かってきたということが多いです。なので、細かい内容を教えるのではなくて、まさに非専門家に対してAIというのは、こう動いているのですよというのをふわっとまるっとお教えします。データを食わせて、学習させて、モデルができますみたいなのを、線形代数とか、確率の内容に関して似たようなたとえ話を用いて話すことがありますね。

みんな実際実務でやられている方々なので、彼らもそのAIを使ってデータ分析してという結果をお客さんに報告しなければならないので、ある程度分かった上で話さなければいけないのだけれども、すばっと抜けている部分を、説明に困らない範囲でも、丁寧にお伝えしておくという感じですかね。

里：ありがとうございます。

参加者からの座談会での感想

里：では、最後に皆さんにひと言ずついただいて、この座談会、終わりにしたいと思います。

先ほどと逆順にしましょう。石井先生からお願いします。

石井：今回、座談会を企画させていただき、「ビッグデータのデータサイエンス」というテーマで、コロナ禍とか、地球温暖化とか、豪雨とか、最近まわりの変化がすごく激しくて、そんな中で自分がどうやって仕事を見つけていくとか、今後どうやって食べていくとか、そういう

ことを考えるいい機会かなと思っています。今日の企画を通して、皆さんの考える材料を提供することができたとしたら幸いだなと思っています。

里：ありがとうございます。

では、牧山さん、お願いします。

牧山：いろいろとお話が聞けて楽しかったです。ありがとうございます。特に大学教育でのAI人材の育成というのは非常に期待していて、今だと機械学習の研究をやっていたみたいなのが企業に来るのですが、ミスマッチになってしまうこともあり、そういう人だけではなくて、サービスに興味を持っていたかつAIの基本的なところをちゃんと学んでいる人たちも来てくれると嬉しいなと思っています。オープンデータの話も面白かったです。みんな同じことに困っているのだなと（笑）。

里：ありがとうございます。

飯尾先生、お願いします。

飯尾：昔、90年代のころというのはITそれ自身の研究で飯が食えたのですよ。それがWindows 95みたいなのが、わあーっと広がってきて、それでITのコモディティ化と言われて、ITそのものの研究だと飯が食えなくなってしまったのですよね。そういうのを経験しているので、なんかそのアナロジーではないのですけれども、最近、そのAIのコモディティ化というのですかね、AIそのものの研究ではなくて、AIをどんなところに応用していくんだというところにどんどんシフトしているではないですか。だからそういう観点で見ていると、結構、いろんなことができて楽しいです（笑）。

最近、変わったところだと、高校生の異文化交流の教育の手伝いなんかもやっていて、本当に全然畑違いのことをITとか、AIとかの支援で参加するような、そんなプロジェクトにもかわり始めているので、やることがどんどんどんどん広がって行って、今、本当に楽しいですね。

里：ありがとうございます。

では、安部さん、お願いします。

安部：そうですね、今の飯尾先生の話にあったみたいな、元々ITできるだけで食えるというところが、これがなくなって、今度はAIが似たような状況になっているというのはまさにビジネスサイドでも似たようなことが起こっているかなというのはすごく感じますね。そんな中で、今のこのリモートワークでしかできないような状況とかを考えると、どういう知識を持って、どういうところに応用していくかという、まさに応用力みたいなところが、人材の価値につながっているのかなという感じがしますね。

逆にいうと、それさえきちんと持っていれば、この状況であっても、生きていけるのではないかなという気はしていますね。とはいえ、私自身がそんな上手くやっているのかという話ではないのですけれども、そういうのを体現するためにも今作っているサービスを成功させて、皆さん

に使っていただきたいなという気はしていますね。データの活用みたいところで、データの流通というのが大事だなというのを私は今思っているところなので、そういうところから何か、支援していけたらなというふうに思っています。

里：ありがとうございます。

では、高柳さん、お願いします。

高柳：私も牧山さんとかの話に近いのですが、非専門家がITを使い始めていて、それでデータ分析も意外とみんなやればできるじゃんという状況にはなっている気がするので、もう少し大学と産業の連携を頑張る。まあ、インターンを増やしすぎると、今度は勉強する時間、学生としての学問をする時間がなくなるので、問題だなとは思いますが、もう少し行き来の自由度を高めていければ、もうちょっと良くなっていくのかな、全体的にデータ活用とか、日本の産業とか、社会とかも全部そうですけれども、良くなっていくのかなというのを、今日話して、ふと思いましたという感想です。

里：はい。ありがとうございました。

安部：里さんから締め言葉はないのですか、里さんの感想なりを（笑）。

里：いや、全然、考えていなかったです（笑）。

石井：ちょうど時間になりました。皆様、本日はお忙しい中座談会にご協力いただきありがとうございました。

脚注

☆1 <https://xtech.nikkei.com/atcl/nxt/news/18/10691/>

☆2 数理・データサイエンス・AI教育プログラム認定制度（リテラシーレベル） | 文部科学省

https://www.mext.go.jp/a_menu/koutou/suuri_datascience_ai/00002.htm

☆3 数理・データサイエンス・AI（応用基礎レベル）モデルカリキュラム～AI×データ活用の実践～ http://www.mi.u-tokyo.ac.jp/consortium/model_ouyoukiso.html

☆4 モデルカリキュラム（応用基礎レベル）案に関する意見募集 http://www.mi.u-tokyo.ac.jp/consortium/mc_ouyoukiso.html

☆5 データサイエンティスト スキルチェックリスト ver3.01

https://www.datascientist.or.jp/common/docs/skillcheck_ver3.00.pdf