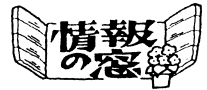


2018年春季シンポジウムルポ（第78回）



田中 未来（統計数理研究所）

2018年3月14日（水）に東海大学高輪キャンパスにおいて東海大学の共催のもと2018年春季シンポジウムが開催された。東京は年明けから厳しい寒さが続いていたが、この日は非常に暖かく天気にも恵まれ、およそ160名の参加者が一同に会した。

今回のシンポジウムでは“データサイエンスの現在（いま）”というテーマのもと、今まさに流行の真っ只中のデータサイエンスに関する5件の講演が集められた。具体的には、産業界からの実際の応用に関する講演が2件と、その基盤をサポートする学術界からの講演が3件とバランスのよい構成となっており、参加者はそれぞれに見聞を広げることができたと思われる。

最初の講演は田島玲氏（ヤフー株式会社Yahoo! JAPAN研究所）による“ログからの機械学習とパイアスへの対処”で、アクセスログからの機械学習に関するヤフー社における取り組みと、現場の問題意識が紹介された。近年のウェブページでは、各ユーザの興味や関心に応じたコンテンツや広告が自動的に配信されることが多い。最適な配信を行うためにアクセスログを解析し、活用するというのは自然な考えだろう。しかし、アクセスログは現在の配信方法に影響されたものであるため、別の配信方法を用いた場合に各ユーザがどのように行動するかを観測することはできず、その際の行動を予測することは根本的に難しい。このような現在の配信方法などがアクセスログに与えるバイアスに対して、統計的因果推論を用いて対処していくつかの試みが紹介された。

ウェブ領域では、複数のユーザに対して別々の配信

方法を適用したり、複数の配信方法を混合して適用したりすることで、大規模なデータを取得することが比較的容易である。これは統計的因果推論が今まで使われてきた医療分野や社会調査分野では見られない特徴である。このような大きく異なる特徴をもつデータに対して既存の手法をそのまま適用することの是非について今後検討する必要があるとの課題意識も紹介された。

次の講演は加藤公一氏（シルバーエッグ・テクノロジー株式会社）による“機械学習のマーケティングへの応用とその注意点”で、シルバーエッグ・テクノロジー社で提供しているレコメンドシステムに関する基本的な手法と実務上の問題意識が紹介された。レコメンド（自動推薦）システムとは、ユーザが興味をもちそうな商品などを自動的に提示する仕組みであり、電子商取引サイトなどで広く用いられている。システムの内部では、協調フィルタリングや行列分解などのORないしは機械学習的な手法が用いられており、これらの手法については現在でも盛んに研究が進められている。

一方で、既存の多くの手法は学習に数時間かけることを前提とするものが多く、リアルタイムでの学習に用いることが難しいという問題意識が提示された。実際、リアルタイムでの学習を行うことにより、新着商品やテレビで紹介された商品を即座に自動的に推薦することなどが可能になるため、リアルタイムでの学習は価値が高い。こういった技術にはオンラインアルゴリズムなどの知見が応用できると考えられ、OR学会員の活躍の場の一つであるように感じられた。

休憩を挟んで3番目の講演は長谷川孝博氏（静岡大



田島氏の講演の様子



加藤氏の講演の様子



長谷川氏の講演の様子



神谷氏の講演の様子

学情報基盤センター)による“データサイエンスを支える学術情報基盤の昨今”で、データサイエンスに不可欠な情報基盤に関するセキュリティやクラウドに関する基本事項に加え、情報システムを素早く展開するための静岡大学の取り組みが紹介された。

データサイエンスでは高い価値の情報を取り扱うことも多いが、それらは常にマルウェアの脅威にさらされている。この講演では、このような情報セキュリティリスクの大きさを価値×脅威×脆弱と考え、唯一制御可能な脆弱に対する対策の重要性が強調された。また、クラウドはこのような情報セキュリティリスクの移転先としても考えられるとの紹介があった。

4番目の講演は神谷直樹氏(統計数理研究所)による“データサイエンティスト育成に関する現状と展望”で、統計教育に関する現状の整理と近年の国内外の動向が紹介された。

この講演では、データサイエンティストがデータから価値を引き出すプロフェッショナルとして定義され、そのためのスキルとして数学・統計学の知識、実装能力、実務経験のすべてが必要となると紹介された。また、これらのスキルをもった人材を育成するための教育についての国際比較が行われ、日本の遅れが指摘された。海外における大学教育では、データサイエンスに関する課程がすでに500以上存在するのに対し、日本では近年になって関連する学部の課程が新設され始めたところである。また、民間企業などによる教育では、海外では教育プログラム修了がその後の雇用に結びつく一方で、日本ではそうっていないのが現状である。また、アメリカでは現状のスキルレベルのデータサイエンティストは足りてきており、今後はよりハイレベルな人材育成が必要になるだろうという予想も紹介された。一方、日本では現状のスキルレベルの人材でさえ不足しているようである。この現状認識に基づく統計数理研究所の人材事業も紹介された。講演の



宇野氏の講演の様子

後の質疑応答では、OR学会におけるデータ解析コンペティションが棟梁レベルのデータサイエンティスト育成に有効だろうという意見もあった。

最後の講演は宇野毅明氏(国立情報学研究所)による“抽象化による理解しやすいデータ解析への展望”で、データ解析における抽象化を利用した手法の紹介とデータ解析に基づく未来の社会について議論がなされた。

この講演では、大規模データ中の個々の情報を抽象化することにより、共通性のある個性を抽出し、データを理解しやすくする方法がいくつか紹介された。また、このような方法でデータが理解しやすくなると、自然にその後の解析の方法が思いつくようになるなど、データ解析がより身近になるだろうという予想がなされた。その結果、政治などのさまざまな分野でデータに基づく議論がなされることが期待される。一方で、現在の日本にはデータ解析の結果に価値を創出する人材が不足しているという問題意識が提示された。これは現実問題をモデル化する能力をもった人材が求められているとも考えられ、OR学会が大いに貢献できるのではないかとと思われる。

以上で十分にお伝えできたかどうか自信のないところではあるが、本シンポジウムは非常に有意義で楽しいものだった。講師のみなさま、司会の中川慶一郎さま、シンポジウム実行委員長の朝日弓未先生をはじめとする実行委員会のみなさまにこの場を借りてお礼申し上げる。