

対談：機械学習の変遷と今後の展望

概要：本原稿は昨今、社会的に大きな注目を集めている機械学習について、慶應義塾大学理工学部櫻井彰人教授との対談内容に基づく記事である。本稿では、はじめに機械学習に用いられる用語の基本的な定義について述べる。次に、機械学習のブームについて、第一次、第二次、そして現在の第三次の3期に分けて、歴史的な背景を含めた変遷を紐解く。最後に、今後の機械学習の展望についてまとめる。

キーワード：人工知能、機械学習、ニューラルネットワーク

1. 経歴

櫻井彰人教授 (対談者)：1975年東京大学工学部計数工学科を卒業後、同大学大学院情報工学研究科において、修士課程を修了され、日立製作所入社那珂工場に就職された。在職中にイリノイ大学 (University of Illinois at Urbana -Champaign) に留学され、修士課程を修了された。1989年日立製作所基礎研究所に配属後、1993年東京大学にて博士 (工学) の学位を取得された。1996年日立製作所中央研究所、1998年北陸先端科学技術大学院大学教授を経て、2001年慶應義塾大学理工学部教授となられ、機械学習、人工神経回路網を専門に研究活動に携わられている。



大竹恒平助教 (インタビュー)：2016年慶應義塾大学理工学研究科開放環境科学専攻後期博士課程修了。現在、中央大学理工学部経営システム工学科助教。マーケティング・サイエンス、ソーシャルメディア情報を活用した経営問題の解決、CSCWに興味を持つ。博士 (工学)



2. 機械学習に関わる諸定義

大竹恒平助教 (以下大竹)：本日は機械学習の変遷と今後の展望について、櫻井彰人先生に色々お伺いしたいと考えております。どうぞ宜しくお願い致します。

櫻井彰人教授 (以下櫻井)：宜しくお願いします。

大竹：2015年のアルファ碁 (AlphaGo) の韓国プロ棋士との対局勝利を皮切りに、ここ数年、人工知能 (Artificial Intelligence) という言葉が様々なメディアで取り上げられ、注目を集めているように感じます。このような状況が後押ししてか、人工知能、機械学習 (Machine Learning)、そしてニューラルネットワーク (Neural Network) という言葉が殆ど同じ様な意味を持つ言葉として用いられている印象を受けます。本日お話しいただくうえで、まずはこの辺りの言葉の整理から始めたいと思います。

櫻井：おっしゃるように、産業界においては今まさに、人工知能ブームが起こっていると思います。これらの言葉の中で概念的に最も広いのは人工知能という言葉ですね。その中に機械学習があって、さらにその中にニューラルネットワーク、最近ではディープニューラルネットワーク (Deep Neural Network) がありますね。

人工知能の始まりは1956年、ジョン・マッカーシーらが主催したダートマス会議と言われています。この会議で人工知能という言葉が用いられ、学術的研究分野として確立しました。

大竹：約2ヶ月間10人の人工知能学者によるブレインストーミングが行われた会議ですね。

櫻井：一方で、機械学習という言葉ができたきっかけは、1959年アーサー・サミュエルが開発したチェッカーのプログラムです。当時のIBM社の技術誌 IBM Journal of Research and Development において、初めて“Machine Learning”という言葉が使われました [1]。これ以降、機械学習という言葉が使われるようになりました。“Machine”という言葉を使ったあたりが、時代背景を表していますね。当時はまだ、computer と他の“Machine”との差異が大きくなかったのかもしれませんが。

大竹：確かに時代背景を表していますね。では、当時の学

†1 慶應義塾大学
†2 中央大学 (連絡先: otake@indsys.chuo-u.ac.jp)

習とはどういった意味合いを持っていたのでしょうか？

櫻井：「学習する」という言葉の定義は時代とともに少しずつ変わってきていますが、概念上の機械が、経験を経て、賢くなり、性能がアップするというのが機械学習の基本的な考え方です。時系列に沿ってデータが増えてくるわけですが、段々と賢くなるという訳ではなく、一步、即ちデータが得られた時に、良い結果が得られるような、賢くなるようなプログラムも機械学習と呼ぶようになってきていると思います。

大竹：現在の学習の意味合いとしては、データ量の時系列的な増加に伴う逐次学習というよりは、データ量が少ない時よりも多い時に結果が良くなるようなアルゴリズムを指しているということでしょうか？

櫻井：はい、そうですね。そして、そこでのアルゴリズムとは何なのかというと、元々の定義からすると何でもいいのです。ですから、サミュエル自身もチェッカーのプログラムにおいて勝率を高くするような盤面評価のアルゴリズムを考えたわけです。

一方で、手持ちのデータを使って学習する際には、実は予測と切っても切れない関係があります。というのも、例えばチェスのプログラムを考えた際には、相手が打つ手を予測するということがあります。他方で、自身が次の手を打った際に最終的に自分が勝つ状況を考えるわけです。これを打ったら最後にどうなるかを計算することは、結果的に予測していることになります。

いつからか、時系列的な予測から、時間・順序という要素とは関係なく、ある意味での未知データに対する予測を行う、というように概念が広がってきたのです。現代では、あるデータセットに対して、学習という動作を行って未知入力データに対する出力値を予測する、ということがすべて機械学習と呼ばれるようになってきているのです。こうしたことから、論理的な関係性はない場合もあるかもしれませんが、現在、人工知能といえ、殆どが機械学習に見えてしまうような時代になってしまっています。

大竹：この辺りが、人工知能イコール機械学習と認識されている要因でしょうか？

櫻井：勿論、各分野で使われている分析方法自身は違います。機械学習の使い方も違いますし、メインとなるアルゴリズムも違います。ただし、機械学習が有用な道具として、数多くの場面で利用される機会が増えています。そのため、機械学習が幅を利かせている様に見え、人工知能と同じものとして認識されているのではないかと思います。とはい

え、それぞれには確かに違いがあると思います。

例えば、私の専門分野ではないのですが、最近では機械翻訳というのが機械学習の牙城となっています。自然言語処理のトップカンファレンスの一つである ACL (Annual Meeting of the Association for Computational Linguistics) の発表では、必ず機械学習がベースにあるという状況になっています。画像処理は言わずもがなですね。とはいえ、分野ごとに特色はあるのです。

機械学習が頭にあるのではなく、ベースに機械学習があるのだと思います。ですので、今後は機械学習を飲み込んだうえで、分野特有のアルゴリズムや工夫を行うという方向になるのではないかと思います。

大竹：すると、機械学習はもはや一つの分野ではなく、様々な領域において、基盤になりつつあるということですね。

櫻井：はい。機械学習を使うことで結果が良くなる、これまでできなかったことができるようになったのが、このような状況を作ったのだと思います。特に人工知能ですと、今まで研究者の知恵でやっていた部分を機械学習に置き換えることで、性能がアップする、誰でもできるようになる、といった結果により、機械学習が基盤として利用されるようになったのだと思います。

大竹：とはいえ、機械学習の本質的な部分というのは大きく変わっていないように思いますが、ここ最近のブームのきっかけは何でしょうか？

櫻井：おっしゃる通り、本質的な部分は何も変わっていません。一つの要因としては、利用できるデータ量が格段に増えたことではないでしょうか。

大竹：これまでソフトウェアの話が中心でしたが、ハードウェアの性能向上によるところも大きいのでしょうか？

櫻井：はい、ハードウェアの話抜きにしては語れませんね。間違いなく、ハードウェアの発展とともに計算スピードが上がり、かつ安価にマシンを購入することが可能となったことが要因として大きいと思います。むしろ、このような状況にならなければ、できなかったことだと思います。

3. 機械学習の変遷

大竹：では、改めまして機械学習の変遷についてお伺いしたいと思います。現在、間違いなく機械学習のブームが来ていると認識しておりますが、このようなブームは過去にもあったのでしょうか？

櫻井: はい, 過去に二度程ブームがあったと考えています.

大竹: では, それぞれのブームについてお話し頂きたいと思います. 合わせまして, 当時の社会情勢や櫻井先生自身がどのような研究を行われていたのかを合わせてお話し頂ければ幸いです.

3.1 第一次

大竹: まず, 第一次ブームについてお話し頂きたいと思います.

櫻井: 歴史的に捉えやすいのは, 人工知能というよりもニューラルネットワークの方ですので, ニューラルネットワークを中心にお話ししたいと思います.

第一次のブームはニューロン素子1個の学習によります. 1960年代爆発的なブームとなり世間の注目を集めました. しかしその後, 1969年にミンスキーとパパートの著書 *Perceptrons* [2] において, 線形分離不可能なパターンを識別できないことが指摘され, それ以上の進展もなくなり第一次ブームは終焉となります.

私は, 当時のブームには直接関与してはいたわけではありませんが, 機械学習分野ではないのですが, 精力的に閾値素子の研究をされていた室賀三郎先生に, 1979年頃お話を伺いました. 当時, ニューロンをモデル化した閾値素子を使うことで, 計算機に用いる真空管の信頼性の低さがカバーできるといったことや, 人間の学習メカニズムが真似できるといったことについて, 確かにブームが起こっていたそうです. 両者が混在してはありましたが, とにかく社会からの期待が大きかったです.

当時, 人間の学習メカニズム, 脳の可塑性の要因はどこにあるか, というところを探り, それはニューロン間の結合の可塑性であることが知られておりました. 脳の可塑性を説明したいというのが研究者の強い要望でしたので, 様々な研究が行われておりました. その中で, パーセプトロンの学習アルゴリズムが, 非常に上手くいったため大きな注目を集めました. しかし, 先に話しましたミンスキーとパパートにより, パーセプトロンはあまり一般化できないことが指摘されました. これにより, 研究が下火となってしまいました.

大竹: 1960年代に人気を博した第一次ブームは, 1969年のミンスキーとパパートの指摘により, 終焉を迎えたのですね.

3.2 第二次

大竹: それでは次に, 第二次ブームについてお話し頂きたいと思います.

櫻井: 第二次のブームは, 1986年並列分散処理論 (PDPモデル) が登場した頃からスタートしました [3]. なお, そのころの機械学習は, 記号表現の獲得を目指す学習が多かったのです. これは, 論理的推論を可能とするため, 研究者の考えを入れやすいという大きなメリットがあります. それと並行して, ニューラルネットワークの研究が台頭してきたわけですが, 言い方によっては, 記号表現の学習で大きな成果が出なかったため, ニューラルネットワークの研究に関心が移ったともいえますが, この辺りは定かではありません. なぜならば, 両コミュニティは殆ど異なる研究者によって形成されていたからです. 多層のニューラルネットワークの学習ができれば, 一般社会に存在する多くの問題が解けそうだと, という期待が生まれたことが当時のブームの始まりだったと思います.

人工知能の素晴らしさを喧伝するのに使われた例として, 積み木の世界での会話, というのがあります. ある仮想的な空間に, 色のついた四角や三角といった積み木が沢山あります. 人間が, 「赤い三角をつかめ」「つかんだ赤い三角を青の立方体に置け」といった指示を出し, その指示に従って仮想的なロボットアームが命令を実行するというシステムです. これは, 記号システムの有用性を示した良い例です. しかし, ここから実際にクレーンを使ってブロックをつかみ上げて, といった実世界で再現することは容易にはできませんでした. この他にも, 言語学習を記号表現として捉えた研究も非常に活発に行われていましたが, 同様に, 実世界での実現は容易ではありませんでした. 当時は機械学習というのは非常に限定されたコミュニティで研究に限られていたといえます.

一方で, ニューラルネットワークが発展した一番の要因は, 学習に誤差逆伝播法 (バックプロパゲーション) を利用したことだと思います. バックプロパゲーション自体は, その時に突然使われ始めた訳ではありません. 古いものですと, 1967年に, 甘利俊一先生が論文 [4] を執筆されています. 余談ですが, 私は修士課程が伊理正夫先生の研究室だったのですが, 当時甘利先生がよくいらっしゃって, 毎日のように碁を打ってらっしゃいましたね. 誤差逆伝播法のアルゴリズム自体は非常に単純ですので, 甘利先生の研究室では, 卒業論文や修士論文で扱っていました. その際に, とにかく時間がかかるという学習スピードに関する問題と, 魅力的な学習課題がない, つまり適切なデータがないという問題がありました.

では, なぜ PDP (Parallel Distributed Processing) の登場によりブームが起こったかという点, 明らかに理由は2つで, データとハードウェアによるものです.

まずハードウェアに関しては, 当時コンピュータが急激に進歩しました. そのため, 最急降下法のスピードが, 大

大きく変わりました。また、ハードウェアの価格が下がったというのも大きな理由だと思います。研究室で気楽にコンピュータを購入できるようになりました。当時はまだ何もわかっていない時代でしたので、初期条件の設定や学習パラメータの設定など、試行錯誤をしながら研究を進める必要がありました。結果が出るまでに多くの時間がかかることは、研究者にとって非常にネガティブな要素です。コンピュータの性能の向上は、試行錯誤の回数を増やすことができるようになり、研究が進展する要因となりました。

大竹：なるほど、確かに研究への参加者が多くなれば研究は発展しますし、分析を実行する環境が整った、というのは非常に大きなインパクトがありそうに感じます。

櫻井：第一次ブームの際には、ハードウェアの価格が物凄く高かったので、参加者がそこまで多くありませんでした。第二次ブームが起きたのは、ハードウェアが速く安くなりだれでも参加できるようになった、というのが大きいでしょう。最急降下法自体は、本当に単純なアルゴリズムですので、大学1年生でも十分にプログラムできます。ブームの最初のころには、卒論や修論かと思える論文が沢山出てきていました。そのくらい新しい試みが溢れていたのです。

もう一つの、データに関しては、突如データが手に入るわけではありませので、苦勞して作っていたと思います。例えば、NetTalk用の音声データや画像認識用のデータは、今となっては手に入り易くなりましたが、昔は苦勞して作成していました。色々な学習に利用可能なデータが揃い、ハードウェア環境が整いましたので、ビジネスへの応用の可能性も出てきました。もう少し後になりますが、ファイナンスデータを用いた学習も行われるようになりました。私が記憶しているのは、1991年のIJCNN (International Joint Conference on Neural Networks) のワークショップに参加した際、ファイナンスのワークショップの大きな会場に溢れんばかりの人が参加されていたことです。データが比較的取得しやすく、予測を行いたいというニーズがあったこと、さらには統計的モデルのように綺麗なモデルではなかなか上手くいかない、何でもいからとにかく良い結果が欲しい、という場面でニューラルネットワークをやってみようという視点で人が多く参入したのだと思います。

大竹：ここまでをまとめますと、第2次ブームのきっかけは、データが比較的揃うようになった、ハードウェアが圧倒的に早くなった、単純なアルゴリズムでも良い結果を得ることが出来た、これらの3点によるところが大きいようですね。

櫻井：はい、そうですね。PDPの本の中身を見ると、フィ

ードフォワードのニューラルネットワークに加えて様々なタイプのニューラルネットワークが検討されています。しかし、それ以外に、様々なアプリケーションが山ほど紹介されています。これを読んだ研究者が、自身の領域に応用してみたいと思うことは不思議ではありませんね。そういった意味で、PDPの存在は大きかったと思います。

この辺りから、機械学習の意味するものが変わってきます。先ほどお話ししたように、従来の記号表現の機械学習とニューラルネットワークの学習とは大きく異なるのですが、外から見ると、機械が学習して賢くなる、データを貰って予測をするという、似た機能を持っています。大きく異なるのは、記号を使った機械学習に対して、ニューラルネットワークは数値を使っている点です。

大竹：確かに、記号表現と数値表現はかなり大きな違いですね。

櫻井：はい、実際に、この違いは後々までかなり効いてきます。

大竹：では、第二次ブームの終焉は、どのようにして起こったのでしょうか？

櫻井：大きく2つの原因があると私は考えています。

一つは、発展の要因の一つがハードウェアの急速な進歩であった点に起因すると思います。ムーアの法則に代表されるように、ハードウェアの性能は漸増的に、コンスタントに上がっていきました。研究というものもコンスタントに進んでいってしかるべきものですので、常識的に考えた際には、どこかでブームが来るというのはおかしな話なのです。ところが、先ほどもお話しした通り、この分野の研究はトライアンドエラーが必要不可欠です。トライアンドエラーをするか否かという研究者の判断が関係してきます。研究者はトライアンドエラーを繰り返していき、ある時点で結果が出ないからこれはもうできないという判断をしているわけです。実際はさらにトライアンドエラーを繰り返せばできるかもしれませんが、それはその時点では分からないのです。ある所までやってもできなければ、できないと結論付けます。そうすると、周りの研究者もできないという認識を持ち、そのまま暫く研究が停滞します。一方で、ハードウェアはコンスタントに進歩していますから、どこかの時点で誰かができることになるのです。逆に言えば、できることが分かるまでは、本当はできているのにやらない、という状況が起こるのです。どこかの時点でできることが分かれば、数多くの研究者が参加してきますし、一気に研究が進展するわけです。そうすると、その時のハードウェアの能力を使い切るまで研究が行われます。より難し

い問題を解こうとし、例えば、ネットワークのノード数を増やします。しかし当然パラメータが増えるので学習スピードは大幅に低下します。ある程度まで大きいネットワークを試みるようになると、先ほどと同様、その時のハードウェアのリソースではもうこれ以上はできないという判断が下されるのです。現時点でのハードウェアのリソースでできることをやり尽くしてしまうと、新たな研究のトピックが生まれず詰ります。人間は定常状態よりも、変化状態を好む性質があるようで、研究が進まなくなると、その分野への注目が減ります。そうすることで、新しく入ってくる研究者も減る、そのため研究が進まなくなる、というのが、第二次ブームの終焉の一つの要因であると考えています。

もう一つの理由としては、これは研究者コミュニティの間のお話なのですが、ニューラルネットワークの性質にあると思います。ニューラルネットワーク研究はやったらできた、という部分もあり、それがどうしてできたか、ということが良くわかっていない部分があります。やれることが増えている際には、こういったネガティブな部分はあまり目立たないのですが、新たな研究成果の差分が小さくなってくると、批判的な目は当然厳しくなってきます。たまたま上手くいったのではないかと今回のデータ特有の話ではないかとといった疑問を払拭できず、結局は一般性がないという判断に至るのです。

また当時、SVM (Support Vector Machine) などの、理論的なバックグラウンドを持ち、汎化能力が示された手法が作られました。そうすると、関心はそちらに向きますね。一見すると中間層1層のニューラルネットワークと同じ様に見えるのですが、学習時間が短く、判別能力もニューラルネットワークが上手くいった際と同程度となっていましたので、ニューラルネットワークから SVM へと関心が移っていきました。

大竹:ニューラルネットワークの分が少し悪くなってきた、というところでしょうか。

櫻井:そうですね。それに加えて、機械学習的にみると、現在のデータマイニングが、1980年代の終わりに出てくるのです(Knowledge Discovery in Databases)。私が1988年に日立の基礎研究所に移った際には、ニューラルネットワークと遺伝的アルゴリズムが新しいテーマとして脚光を浴びていました。このころ、データマイニングという言葉が脚光を浴び始めました。特にビジネスにおいては、データはあるけど、どう分析すればいいかわからないという状況も多々ありました。このような状況への打開策として、アグラワールらが開発した、トランザクションデータからの知見獲得を可能としたアソシエーションルールマイニングなど、

従来の機械学習ともニューラルネットワークとも違うアプローチの与えた影響は、非常に大きかったです [5]。このころ、Discovery という言葉が非常に流行りました。

少し話をまとめますと、この頃世間的に注目を浴びたのはデータマイニングで、従来の記号系の機械学習は、少し注目を浴びる機会が減っていきました。ニューラルネットワーク系は、ブームになった後暫く研究が停滞します。その他の方法としては、例えば遺伝的アルゴリズム (Genetic Algorithm) などが注目を集めていました。最初に機械学習は段々賢くなっていくという話をしましたが、そこから段々とデータからあるモデルを作成する、作成したモデルを用いて予測を行うという風にシフトしてきました。これらがまとめて意識されるようになったのは、この頃だと思います。

ビジネス的なインパクトとしては、データマイニングの登場は大きかったと思います。機械学習やニューラルネットワークのコミュニティから、データマイニングの領域に研究者が参入しました。ニューラルネットワークの観点からは逆風でしたが、データ解析という観点では、大変実りの多い時代のスタートになりました。

これは私の個人的な体験ですが、当時例えば日立では大型コンピュータを売っていました。しかしながら、高価であり、ワークステーションのような小型コンピュータの台頭により、大型コンピュータが売れにくくなっていました。一方で、アメリカではデータマイニングが大流行していましたので、計算能力で勝る大型計算機を売るチャンスであるとも考えていました。これは、計算機で大量のデータを処理することへのビジネスにおけるニーズがあり、これをメーカーとして積極的にサポートすることを始めるきっかけであったと思います。

3.3 第三次

大竹:それでは、今日に迫る第三次ブームについてお話し頂きたいと思います。

櫻井:第三次ブームのトリガーを考えると、やはりインターネットの登場だと思います。インターネットを利用することで、様々なデータにアクセスすることができ、データ量が増えました。当初は接続コストが高く、機能も非常に限定的でしたので、とても今のような状況になるとは想定していませんでした。ただ、このころからデータの蓄積は始まります。例えば、ニュースグループという、今でいう掲示板のようなものなのですが、常時アクセスではなくメールにてニュースを送るサービスがありました。また、Netscape 社のブラウザを利用することで、インターネットの先にある画像を表示できるようになりました。テキストや画像といったデータが蓄積され始めたのが、この時代か

らであると思います。特に画像の送受信は、通信コストの低下や、データ通信量の増加といった、データ蓄積に対して正の循環関係を形成しました。こうして蓄積されたデータをどう使うか、考えるようにもなりました。

大竹：まさに、ビジネスシーンにおけるデータマイニングの始まるきっかけと同じ状況が、インターネット上のデータを蓄積するうえで、でき上がったということですね。

櫻井：はい、ここで機械学習を使おうという流れが形成されます。当時、機械学習で扱えるデータ量はまだまだ小さいものでした。これは単純にハードウェアの制限によるものでした。

今の機械学習ブームのきっかけは、機械学習の部分ではなく、「ビッグデータ」がキーワードであったことは間違いありません。インターネットを通じたデータ蓄積、データ蓄積量の拡大という背景のもと、ビッグデータの活用に注目が集まりました。手法としては、機械学習から発展してきた手法や、データマイニング系の手法が主として使われています。先ほどもお話した通り、元々データマイニングというのは、機械学習のコミュニティにいた人のうち、アプリケーションに目を向けた人たちが作った言葉ですが、記号学習に拘らず、数値学習を積極的に用いた人たちです。彼らもハードウェアの発達に従って、ある程度大きなデータが処理できるようになりました。それに従い、所謂機械学習の人たちも、データマイニングの人たちも、同じようなテクニックを使うようになりました。分かれていたものがいつの間にか一緒になったのです。データマイニングがスタートしてビッグデータという言葉ができたのですが、今では殆ど区別がないようにすら感じます。

大竹：最近では、データサイエンティストという言葉ができていますね。

櫻井：はい、データ分析がごく普通に行われるような時代になりましたね。

ニューラルネットワークに話を戻しますと、多層ニューラルネットワークの学習アルゴリズムは以前からあったのですが、勾配消失問題のせいで実際の問題で利用することが出来ませんでした。アルゴリズムの性質から考えれば、消えて当たり前ではありますけどね。データによっては、中間層が2, 3層程度であればできたのですが、一般的には難しかったのです。ブレイクスルーのきっかけになったのは、勾配を使わないで下層からオートエンコーダを重ねていく方法を考えた、ヒントンの研究でしょう [6]。できるといわれるとやってみようという研究者が参入し、様々な方法が考えられてきます。従来の誤差逆伝播法を用いた場

合でも工夫することで利用可能である、所謂加速方法に気付いた人たちもいますね。当時、私も含めて中間層を2層にして学習できるかわからないよりは、中間層1層で学習すれば十分ではないか、という意識がありました。ところが多層ができてみると、1層ではできなかったことができるようになるのです。これは大きいですね。できなかったことができるようになると、当然もっと別のことをやってみる人たちがでてくるのです。それにより、研究は加速するのです。

現状では、end-to-end で学習ができるようになっていますが、これは非常に大きいです。ヒントンのオートエンコーダを重ねる方法では、教師なし学習をやっていることになり、教師なし学習よりは、教師付き学習をやったほうが、課題に特化した特徴量の抽出を行うようになりますので、教師なし学習を積み重ねて最後に教師付き学習をするよりは性能が良くなります。ただ、どうして浅いネットワークではだめなのか、については今のところ誰も答えていないように思います。とはいえ、できるようになったことは、他にも使ってみようというモチベーションになります。また、今のディープラーニング系の研究者たちは、GAN (Generative Adversarial Network) に代表されるように、アーキテクチャそのものを変えようとしています。全く新しいファンクションを作ろうという動きもあります。

4. 機械学習の今後の展望

大竹：機械学習の今後の展望について、櫻井先生のご意見をお聞かせ頂きたいと思います。はじめに、現在のブームはいつまで続くとお考えでしょうか？

櫻井：ニューラルネットワークが第一次ブーム、第二次ブームときて、今を第三次のブームとすると、私は、新しい機能、もっと言えば新しくできそうなことが提示されているうちは、ニューラルネットワークは発展していくと思います。ただしこの先、ハードウェアの性能をソフトウェアが食いつぶす、つまりその時のハードウェアでできることをしつくしてしまう時が来る時が来ると思います。新たな適用分野がいつまで出現するのかで、今後何年続くのかが決まると思っています。新しい分野が現実社会で役に立つようになるのは遥に先になることもあるので、役に立たないというネガティブな意見がでるかもしれませんが、新しい分野で新しい結果が出れば新しい研究者は入ってきます。新しい結果が色々出続ければ、運次第ではありますがどれかは実際に使えるようになります。ビジネスでは、使えるかなと思いつきながら着目し、資金を提供することで、ベンチャー企業が台頭してきます。これは循環なので一度回り出したら暫く回ると思います。今のところ、うまく回っているという印象を持っています。

大竹：確かに、最近ではディープラーニング系のベンチャー企業も増えていますね。

櫻井：一方で、第二次と第三次ブームの間に広がった、SVMなどの手法との関係のみをみましょう。現在は、やったら上手くいったが先行している状況であり、どうして上手くいったかがいまだに良くわからない部分もあります。これは、今まさに第二次ブームの時と同じです。理論に強い研究者は、今の新しく出てきているアーキテクチャがどうして上手くいくかの検討や、しっかりした理論的基盤に立つ代替物を発明してくれるかもしれません。それが登場した際には、第二次ブームと同様に、理論的な背景を有する手法が選好されます。つまり、進歩の歩みが緩やかになり、代替技術が登場したことがきっかけで一気にブームが陰り、代替技術が広く使われるようになる、ということは考えられます。ビジネスアプリケーションとしては、ニューラルネットワークの勢いは陰るかもしれませんが、データ解析の発展はコンスタントに続いていくと思います。

大竹：私自身、共同研究などで民間の企業の方と交流する機会が多いのですが、全体の印象としては、データはあるけどどうしていいかわからない、という問題意識を持っているように感じます。その問題解決方法として、人工知能、機械学習、ニューラルネットワークに対して過度の期待を感じることがあります。これはメディアの影響もあると思うのですが、今のブームをどのように感じられておりますでしょうか？

櫻井：私を含めて第二次ブームを知っている人は殆ど共通していると思いますが、今のブームに関してはかなり冷めた目で見ています。結局は、第一次、第二次と同じことを繰り返しているというように思います。もちろん、興味を持って仕事を持ってきてくれる人は大歓迎です。誤解している部分に関しては、誤解を解消するように気を付けています。特に難しいデータを持ってきてくれると嬉しいですね。

大竹：はい、研究室ではお目にかかれないデータが殆どですね。

櫻井：ただ、多くは誤解して来られるので、人工知能と機械学習の違いでしたり、機械学習の本質はそんなに新しいものではないことを伝えます。ただし、幾つもの技法が基盤的にもしっかりとしており、データの規則性を見つけて何らかの結果を出すことができることについてはお話ししています。ただ、これらは決して今メディアで騒がれてい

るようなキラキラしたものではなく、至極まっとうな結果しかでないというのは念押ししますね。あまりネガティブなことをいうのも良くないですけどね。私がお付き合いするのは比較的、技術職の方が多いので、既にわかっている場合も多いです。

大竹：最近研究として取り組まれていることで、難しいと感じたものはありますか？

櫻井：今私が難しいと感じているのは、多人数の経験を集めたデータから得られるモデルと、個々の人が自分で自分の経験から築き上げたモデルで予測が異なった場合、どちらが正しいのかという問題です。例えば、タクシーを拾う人の予測があります。ベテランの運転手よりも、多くのタクシー乗車の記録を集めたデータに基づきコンピュータが予測する方が正しいことが知られています。今私が取り組んでいることで結論が出ていないものの一つですが、例えばトンネルでコンクリートの状態を調べるということを考えます。コンクリートは経年変化でひびが入ります。そのひびが大きくなっていくと壊れてしまいます。勿論壊れる前、もっと言えば直前ではなくもっと前に致命的になるひびの存在を知って修理をしたいのです。あるひびが見つかったときに、このひびが将来崩落につながるかどうかを予測したい。それが10年後なのか20年後なのかを予測したいのです。なぜかという、そういう予測ができるベテランは、今後減る一方なのですが、予測しなければいけないトンネルは沢山あります。単純に考えると、ひびを写真に撮って、画像使って診断すればいいと思うのですが、実は簡単ではないのです。なぜかという、ひびの中には、徐々に大きくなるひびと、最初の収縮によりできる全く問題のないひびがあり、今あるひびが将来崩落につながるようなひびに発達するかどうかの判断は、そう簡単ではありません。素人目には全然わかりません。ところがプロになると分かるというのです。そこで、画像からひびをみつけ、計測したひびの長さや幅を使った予測システムを作ろうとしたら、その教師データに困るのです。なぜかという、このひびは危ない、というのはその時点での専門家の判断であり、本当にそのひびが発達して崩落したかどうかを見届けた訳ではないからです。基本的には先輩から教わることで判断しているのです。ですので、エキスパートが持っている知識が、学習データの正解ラベルとして扱っていいのかが分からないのです。経験上かなり確からしいというのは知られていますが、実際に壊れるところまで見た人はごくわずかなのです。一方で、コンクリートの性質を使ってシミュレーションすることができますので、シミュレータの答えというのがあります。人間のエキスパートが言っていることが正しいのか、シミュレータが正しいのか、それ

ともそういったこととは関係なしに存在しているのか、それが正しいのでしょうか。正解率を適切に定義するのが難しいのです。このように、正解ラベルをどこまで信用しているのかわからない場合にどうすればいいのか、という問題は難しいと感じています。こうした問題は人の命にかかわることなので、繊細であり大変難しい問題であると考えております。これは、自動運転を考える際にも表れますね。

大竹：それでは最後になりますが、今後、機械学習で注目すべき新しい分野といったものはありますか？

櫻井：全く新しい分野というと、中々思いつきませんね。

大竹：それ位、どこにでもベースの技術になっているということですね。

櫻井：はい、データが取得できるようなところでは基本的にはできると思います。人間との違いが重要になる場合というのが良くあると思います。人間の知能は、基本的には自分が経験したものからしか得ることができません。勿論、本などから他者の体験を自分の体験とすることができる、といいますが限られると思います。どの分野でもそうですが、他者の体験も自分の体験も含めて、ものを書いて渡す、といったことは殆どないです。エキスパートシステムの考え方として、専門家の知識は専門家しか持っておらず、ナレッジエンジニアが一緒懸命取り出すというのは、真実だと思います。これは、永遠に変わらないと思います。この

ようなときに、個々が持っている知識・経験その記述ではなく、経験したことそのものを数値化することが必要であると思います。数人のデータではなく、百人、千人が集まることによって、人間よりも優れたモノができる可能性があります。こういう分野であれば、何でも適用できると思います。

大竹：大変興味深いお話し、どうも有難うございました。

櫻井：有難うございました。

参考文献

- [1] Samuel, A. L., "Some Studies in Machine Learning Using the Game of Checkers," *IBM Journal of Research and Development*, Vol. 3, Issue. 3, pp. 210-229 (1959).
- [2] Minsky, M. and Papert, S., *Perceptrons: an Introduction to Computational Geometry*, MIT Press (1969).
- [3] Rumelhart, D. E., McClelland, J. L. and PDP Research Group, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Foundations*, MIT Press (1986).
- [4] Amari, S., "A Theory of Adaptive Pattern Classifiers," *IEEE Transactions on Electronic Computers*, Vol. EC-16, Issue. 3, pp. 299-307 (1967).
- [5] Agrawal, R., Imieliński, T. and Awami, S., "Mining Association Rules between Sets of Items in Large Databases," *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data*, pp. 207-216 (1993).
- [6] Hinton, G. E and Salakhutdinov, R. R., "Reducing the Dimensionality of Data with Neural Networks," *Science*, Vol. 313, pp. 504-507 (2006).