

[2019年 秋講演会]

機械学習・ディープラーニングの進展・普及と課題

慶応義塾大学 理工学部 教授, 放送大学客員教授

山口 高平 氏

この記事は、情報システム学会 秋講演会（2019年10月16日）における講演の口述内容をまとめたものです。

■はじめに 第2次 AI ブーム

皆さん、こんばんは。今、ご紹介いただきましたように、私は、AIについて学部4年生の時から、知識ベース推論、今日のテーマでもある機械学習、最近ではAIロボットのような研究を40年以上続けてきました。AIには流行り廃りがあります。1980年代の第2次AIブームの時は、第5世代コンピュータ、エキスパートシステムと言う言葉がもてはやされました。私は、ほぼ毎週どこかの企業でエキスパートシステム構築法に関するセミナーをやっていた記憶があります。でも、専門家の持っている知識をコンピュータ内部で表現することは非常にコストがかかりました。専門家が言葉で説明できる知識を形式知と言いますが、その形式知以外の暗黙知を大量に持っていて、その暗黙知をすべて聞き出して、形式知として記述することは不可能でした。そのため、形式知だけでエキスパートシステムを開発しても高い性能は見込めませんでしたので、性能に限界が見え始め、1990年代には関心が寄せられなくなりました。

その後に登場したのが今日の「機械学習」というキーワードです。データを与えて、データに内在するパターン（知識）を自動的にコンピュータが学習する「機械学習」

が登場し、第3次AIブームでの中心的技術になっています。ただ、機械学習という言葉は様々な意味があり、技術用語として使う場合は、あとで説明しますが、ある特定のアルゴリズム群を機械学習と呼び、今日のタイトルの二つ目にあるこのディープラーニングとは差別化します。機械学習といえば、統計数理に基づく機械学習を意味します。しかし、メディア的には、ディープラーニングを含めて機械学習という用語を使う場合が多いので、注意が必要です。

また、機械学習とは別に、データマイニング、データサイエンスという用語もあります。これらは機械学習と同じ意味と思っておられる方もいると思いますが、専門的には違います。この点についても後でお話します。

機械学習、データマイニング、データサイエンス、ニューラルネットワーク、ディープラーニング、様々なAI用語が世の中で飛び交り、『データは21世紀のビジネスの考える新しい石油である』というような誇張された標語も生まれ、これについても後で言及しますが、データと石油の対比はミスマッチです。一つの情報資源ではありますが、石油ほどきれいなものではないことを後で説明します。

「データを保有していれば AI で何かできる」という考えが上司から技術部門に下りてきて、そこから私どもに相談が来るケースがあります。1-2 か月に一度程度ですが、「企業の技術部門がこういう相談をしにきたのですが、どうしたらいいのでしょうか？」と聞かれるので、「まずは上司に AI を勉強させるべきだ」というように回答しています。AI で解決できる場合もあれば、できない場合もあります。それを見極める能力を身に着けることが重要です。そのためには、詳細なアルゴリズムレベルで AI を理解する必要はありません。AI にも得手、不得手がありますので、ポイントを押さえて理解しておけば、これは AI でやってもなかなか難しいよね、と判断できるようになります。ユーザーにはその判断力が求められます。

今年 6 月に内閣府から、AI 人材 25 万人育成計画が打ち出されました。8 月末に具体案を各大学に通達すると言われていましたが、まだ何も来ていません。AI の専門家からすると、AI 人材の前にまず AI 教員を育成するべきだと思います。教える人がいないのに、どうやって 25 万人もの AI 人材を育成できるんだ、という事です。今日、最新の学会誌の巻頭言を資料として配布させていただきましたけれども、AI における中国の台頭は目覚ましいものがあります。

3 年前は、AI で大したことがなかった中国が、2030 年には AI で世界一になると、習近平国家主席はそのように宣言し、現在、巨額の研究開発費が投入されて、AI 社会実装が進んでいます。新しい AI が中国から生まれているかと聞かれたら、それは NO だと思いますが、社会実装についてはものす

ごい勢いで進んでいます。例えば、深圳（しんせん）の自動運転の実験は目を見張るものがあります。このように世界では、国策としての AI が進んでいます。それで、我が国の内閣府も AI 人材育成を提言し始め、AI 教員をまず育成するべきですが、そんな悠長なことは言うてはおられないので、AI 教員育成と AI 人材育成を同時並行で進めていくしかないという状況です。

先程、説明しましたように、2 回目の AI ブームのエキスパートシステムはかなり期待されましたが、ブームが去った後は、残念ながら、社会に残したものは大きくありません。しかしながら、現在の第 3 次 AI ブームで出てきた機械学習、ディープラーニングを中心とするシステムは、今後、インターネットやデータベースと並んで、情報インフラになるであろうという意見が多くなっています。でも、AI にも得手不得手がありますから、まずその付き合い方を学ぶ必要がありますので、今日の講演もそのあたりを中心に話します。大学では数式をけっこう使いますが、今日は、一切数式は使いませんので、AI のこの技術はこういう特徴があるのか、そういうレベルで理解していただければと思います。

■機械学習の歴史

機械学習の歴史は、けっこう古くて 50 年前に遡ります。アーサー・サミュエル、この人はコンピュータ・チェスの研究者で、「アルゴリズムを与えなくてもコンピュータが自律的に問題解決パターンを見つけて問題解決能力を向上する仕組み」が機械学習であると定義しました。そのあと、教師

有り学習、教師と言うのは答えであり、例えば画像理解の場合、この写真は猫ですよ、この写真は犬ですよ、というように答を与えて学習していく、教師有り学習が研究されました。一方、教師無し学習は、そのような答えを一切与えない機械学習であり、主にクラスタリングに利用されます。クラスタリングとは、答えを与えずに、ある観点からデータの類似性を考え、全体のデータ集合を分割していく方法です。

半教師学習（教師有り学習と教師無し学習の統合）、強化学習も研究されていきます。強化学習は、ロボットの動作学習でよく使われ、例えば、ペットボトルを掴む場合、人間なら簡単に掴めますが、ロボットに掴ませようとすると、アプローチする腕の角度、掴む時の握力時系列変化など、様々な物理パラメータを調整する必要があります。大変な作業になります。私の研究室では今、ロボット喫茶店という実験していますが、ロボットアームに何千回も把持実験をさせないとうまく掴めません。この状況で、掴み方を失敗したら今の動作が悪かったということでマイナス点をつけ、たまたま上手く掴めたら、その掴み方にプラス点を与え、それをベースにしてどんどん発展させていく学習が強化学習となります。成功すればプラス、失敗すればマイナス点をつけて、正しい知識、動作を学習していくことになります。ロボット喫茶店で数千回と言いましたが、工場で実際にピッキングを色々やらせるためには、数十万回の実験が必要になるケースもあり、製造業の人にそれを言うと、そんな数十万回の実験を当社でできないと言われ、強化学習の実践に

はこのような壁があります。

グーグルなどは、何百台、何千台のロボットを並列に走らせて、数週間かけて実験を行います。強化学習は、そういう意味でコンピュータ・ロボット設備を持っている組織が強く、設備がなければ、強化学習の実践は難しいとも言えます。後でも説明しますが、機械学習・ディープラーニングでも同様の壁があります。大規模データを保有し、高速計算機設備を長時間動かせる環境にある組織が、機械学習、ディープラーニングの成果を出す傾向にあります。これって、あまりインテリジェンスを感じない、腕力勝負の AI だなあというイメージがありますよね。大規模データや多くの試行回数を必要としない機械学習の研究も大学で進み始めていますが、まだ、実践ではなかなか追いつけない現状があります。

機械学習の話に戻しましょう。伝統的な教師有り学習として、人が学習結果を理解しやすい特徴をもつ決定木学習があり、現場でも最初によく試されています。また、伝統的な教師無し学習（クラスタリング）として、K-means 法があり、これも現場で最初に試されています。教師有り学習は、色々と発展を遂げ、POSデータでどの商品とどの商品が一緒に買われるか、購買分析によく使われる方法として、相関ルールがあります。また、確率的に考えていく機械学習として、ベイジアンネットがあります。さらに、サポートベクターマシン(SVM)というものが出てきて、単独の機械学習アルゴリズムでは、SVM の学習性能が最高という結果が出ています。また、最近では、Boosting とかランダムフォレストという、

多くの学習結果を総合的に判断する方法論が考え出されています。いくつかの機械学習を走らせて多数決で決めていくのです。実際に機械学習をするのは、決定木学習とかサポートベクターマシンなのですが、複数の機械学習を走らせて（決定木学習でもパラメータを変えると学習方法が変わります）、こういう場面では SVM の学習結果を採用する、こういう場面ではベイジアンネットワークの学習結果を採用する、というような、メタレベルで判断する機械学習法で、現在、ランダムフォレストという方法が良く使われています。

しかしながら、応用面から考えると、コストが大きいのは、機械学習法の実装ではなく、人がデータの特徴量を考えて与えないといけないことです。あとでサッカーデータマイニングに触れますけれど、そこで、一次データだけをそのまま与えて学習できればいいのですが、それは通常難しいのです。事前にあるデータは、そのまま使っても役に立たないと思った方がいいでしょう。こういう見方をするとこのデータにはこういう特徴があるよ、そういう特徴量を一生懸命考えて、それが良ければ良い学習結果が出るし、それが悪ければ機械学習はあまり機能しません。そういう実態があります。ただディープラーニングに比べると、機械学習は小規模データからの学習が可能です。

■第3次ニューラルネットワークブームの背景と現状

一方、ニューラルネットワークは、AI と同じように、今、3 回目のブームが来ています。第3次ニューラルネットワークブーム

ム＝第3次 AI ブームと言えます。

従来の AI ブームでは、1960 年代に探索・推論技術が第1次 AI ブームを牽引し、1980 年代にエキスパートシステムなどの知識処理技術が第2次 AI ブームを牽引したのですが、第3次 AI ブームは、ディープラーニング+機械学習が牽引しています。機械学習は、記号処理 AI ですが、ディープラーニングはニューラルネットワークであり、記号処理とは異なる分野です。

ニューラルネットワークは、1960 年代に単純パーセプトロンと呼ばれる 2 層のニューラルネットワーク、1980 年代に多層パーセプトロンと呼ばれる 3~4 層のニューラルネットワークがブームになりました。機械学習との差異は、特徴量を考える必要はなく、生データをそのまま与えればいいのです。しかしながら、与えたデータにマッチしすぎて新しいデータを判断できないという過学習 (overfitting) 問題、もうひとつは、30 年前、コンピュータはまだ遅かったので、学習時間が長くなるという問題がありました。私も 1980 年代、多層パーセプトロンでプラント故障診断を学習させましたが、学習に一週間位かかるのはザラでした。コンピュータを回し続けてもなかなか学習結果が出ない状況でした。その時、先程説明した SVM が登場し、様相が一気に変わりました。SVMの方が、多層パーセプトロンより学習性能が高く、学習時間も短いし、データも少なくても済み、良いこと尽くしです。端的に言えば、多層パーセプトロンは SVM に敗北したのです。ベンチマークの競争結果をみても、どのデータでも SVM が勝っているわけで、多層パーセプト

ロンを使っても仕方ないねというムードが漂い、第2次ニューラルネットワークブームは終焉を迎えます。

でも、トロント大学のジェフリー・ヒント教授は、長年、ニューラルネットワークの研究をされており、5層以上でも学習時間が長くない方法を考え、2012年頃、8層のニューラルネットワークを実現しました。ちなみに彼は、ユニークな学者で、立ったまま議論するそうで、疲れるから、余計な話しをせず、自然と議論に集中する雰囲気になると聞いたことがあります。そして8層程度以上の多層ニューラルネットワークのことをディープラーニングと呼ぶようになりました。現在、メディアでもこの用語がよく使われるようになり、日本語では深層学習と呼びますが、このディープラーニングが、SVMを代表とする従来の統計確率的機械学習を圧倒するのです。

ただし、先程も指摘しましたように、高性能コンピュータと大規模データが必要です。そういう計算力とデータ量を備えた組織となると、やはりGAFAsのようなデータプラットフォームになります。このため、理論は大学で考えられましたけれど、社会実装は、GAFAsやマイクロソフトなどがリードしているという実態があります。

■ 視覚認識と畳み込み型ニューラルネットワークの全体像

ディープラーニングは、3種類に分けられるのですが、目の代わりになる視覚認識系のディープラーニングとして、畳み込み型ニューラルネットワークがあり、社会実装でも一番よく使われています。

このような犬の画像があったとします。224×224、それぞれのピクセル(ドット)を入力層の各ノードに与え、中間層として畳み込み層とプーリング層が複数回セットし、認識結果を出力層の各ノードに与え、2012年頃、8層程度のニューラルネットワークが実装されました。

少し技術的な話になりますが、畳み込み層というのは例えば、ここで1のあるところが点だとすると、対角線上の特徴を取りたいときは、原データ、先ほどの犬の画像とかの原データに順番に適用していくわけです。この部分とこの対角線の部分のANDをとると答としてこれが出てくるわけです。1と1なら1、ここは0と1だから答はゼロになります。この部分は1だけこちらには0です。だからこの答は0になります。そうすると、このANDをとった結果を4回スライドさせたら5×5の入力データを網羅するので、答が出てきてここでこの値を足すわけです。ある意味で、ここでは対角線という特徴を考えていて、この対角線の特徴が5点満点で、この部分は4点で対角線特徴が大きい、この右上にも4点で特徴が大きい、でも左下には点が低く対角線の特徴は小さいというように、人間からするとこのような理解になります。あとは単にこれらの最大値をとってデータ圧縮しているだけです。

一方、こちらは車の型、車種を認識する畳み込み型ニューラルネットワークです。この画像は小さくて見えないと思いますが、アウディとかベンツとかトヨタなどの色々な車種の写真が認識対象となります。最初の入力層では、この画像の一つの点が一

のノードに対応しており、フィルターを通して対角線をとると、入力層に近い中間層にこのような線が出てきています。さらに進んで、中間層が出力層に近くなると、窓枠とかタイヤとか丸い物とか、図形レベルの特徴量が学習されてきて、最後の出力層で車種が判別されるわけです。

従来の機械学習では、車種判別だから、注目すべき特徴量は窓、タイヤであると、人間が事前に考慮して、それらの特徴量を反映したデータを準備し、SVMなどの機械学習に与える必要がありました。その特徴量が適切でないと、学習性能は出ず、再び、別の特徴量を考案する必要がありました。でも、ディープラーニングでは、大量画像データ、何千万枚とか何億万枚と必要となる場合もありますが、生データを与えれば、そのような特徴量は中間層に自動的に生成されるので、大きな関心が寄せられたわけです。高速コンピュータを長時間実行させる必要はありますが、そのような計算機設備と大規模データがあれば、自動的に学習できるのです。すなわち、人間が考案していた特徴量をディープラーニングは自分で見つけるのです。学習当初は、認識は失敗するのですが、失敗したら、正解になるまで、ノード間のリンクの重みの組み合わせを少しずつ変えるのです。このノード間の組み合わせは非常に多くあるため、重みを少しずつ変えることに長時間かかり、普通の計算機では時間がかかりすぎるので、高速計算機が必要になるのです。

2012年、Googleの猫と呼ばれた畳み込み型ニューラルネットワークは、猫などの動物や人の顔の多数の写真画像に対して、

1000種類位の動物を見分けるディープラーニングで、高い認識精度を達成しました。この結果は、ウォールストリートジャーナルなどの一般紙にも取り上げられ、人の目に代わるようなAIが出てきたというムードが社会全般に広がったわけです。

■イメージネット

イメージネットと呼ばれる画像認識コンテストがあります。この棒グラフは、横軸が年代で、右のほうが古くて最右が2010年、縦軸は画像認識のエラー率です。2011年は、SVMのエラー率は25.8%でした。他の機械学習も適用されましたが、25%よりエラー率は低くなりませんでした。記号的機械学習の性能限界が、エラー率25%だったので。人間が試みても、だまし絵のような写真もあり、多少は間違えてしまい、エラー率は5%程度です。人間は100枚の画像を見せられて5枚間違える、機械学習は100枚のうち25枚間違えるので、人間の眼はコンピュータの眼と比べて、格段に性能が高く、この差は何ともしがたいと言われていました。そして、2012年、ジェフリー・ヒント教授のチームが、イメージネットのコンテストに参加してきて、AlexNetと呼ばれる8層の畳み込み型ニューラルネットワークを適用し、エラー率が16.4%となりました。25.8%から16.4%と、10%程度、エラー率が急激に下がったわけです。そして2013年以降、ディープラーニングの層数が増え、エラー率も下がっていき、2014年Googleが22層のディープラーニングを考案し、エラー率が6.7%となり、人間の眼の性能5%に近づいてきました。

そして、従来から画像処理研究で定評のあるマイクロソフト社が、2015年、ResNetと呼ばれる150層程度の畳み込み型ニューラルネットワークを考案しました。ResNetでは、リンクの重みを変化させる方法だけでは、学習収束に長時間かかるので、リンクをショートカットさせる方法と併用させることを考案し、その結果、エラー率はさらに低下して3.57%となり、人間の眼の性能を超えてしまいました。6.7%から3.57%だから、エラー率が3%下がっただけと思われるかもしれませんが、すでに認識精度が高いレベルで1%精度を向上させることは大変なことです。認識精度が低くければ、エラー率を数%下げることが容易ですが、エラー率が6.7%という認識精度が高いレベルで、さらに3%下げるのは至難の業です。その困難なタスクを、マイクロソフトがResNetと呼ばれる150層のディープラーニングで達成したわけです。

この後、多くの中国企業が参入してきて、ResNetをカスタマイズしたSeNetが開発されました。150層から115層と層数を少なくしたにも関わらず、エラー率がさらに1%程度低下し、2.3%になりました。人間のエラー率は5%でしたので、エラー率が半分以下になったわけで、コンピュータの眼は人間の眼を超えてしまった、といわれるようになったわけです。なおイメージネットのコンテストは、一定の役割を果たしたので、現在は実施されていません。

■大腸ガンの診断

ちょっと、映像をお見せしましょう。これが畳み込み型ニューラルネットワークを

使った大腸ガンを発見するディープラーニングです。国立がん研究センターとNECの共同研究です。NHK Eテレで、東大の松尾豊先生が監修されている番組です。その映像の一部をご覧ください。

<ビデオ再生>

このビデオは大腸がんでしたが、この他にも、胃がんなどにも適用され、がん発見率は98%程度になっています。医師の発見率は80%程度ですので、ディープラーニングにがんの画像診断をしてもらった方が良いでしょう。ただし、画像診断後の手術計画などは、AIはまだまだ医師には及びませんので、画像医療診断はディープラーニング、その結果を受けて、医師が色々と考えるという協働関係の可能性が出てきたと言えます。ただ現在は、法律などの壁があり、現場では普及できていない状況です。

以上まとめますと、現在、100層程度の畳み込み型ニューラルネットワークが、画像認識ツールとして様々な分野で適用されつつあります。

従来、人間が考えていた特徴量をディープラーニングでは中間層で自動生成できます。ただし、ディープラーニングには、大規模データが必要で、高速コンピュータも必要です。そのため、世界レベルでは、GAFAなどのデータプラットフォーマーが強いという現状があります。

■言語処理とディープラーニング

自然言語（人が使う言語）処理のためのディープラーニングの研究は、この1-2年、

急速に進展し、この分野はグーグルの研究がリードしています。画像処理系ではマイクロソフト、言語処理系はグーグルですね。2018年10月、グーグルがBERT (Bidirectional Encoder Representations from Transformers, バート)と呼ばれる自然言語処理系のディープラーニングを発表しました。画像認識と異なり、自然言語処理では、文脈を理解することが重要ですので、従来、時系列データを扱うディープラーニングであるRNN (Recurrent Neural Network, 回帰ニューラルネットワーク)が自然言語処理で使われていたのですが、一方向だけの文脈の処理でしたので性能に限界があり、BERTでは、両方向の文脈処理を実現して、当該文章の前後で、重要な単語(アテンション)を自動的に発見する機構が考案され、自然言語処理の性能が飛躍的に向上しました。その結果、Wikipediaを用いたクイズ解答などでは、人間の能力を超えるケースが出てくるまでの状況になってきました。

2019年8月、マカオで開催された人工知能国際会議でも、BERTの応用に関する発表が相次ぎ、Q&Aシステムへの応用例が多く発表されていました。2019年6月に発表されたXLNetはBERTを改良し、BERTの性能を超えることに成功しており、日々、このような改良が続いています。ディープラーニングの論文は年間1万件、機械学習を含めれば、年間3万件にもなると言われています。とてもすべてフォローできませんが、世界中の多くの研究者が、この研究分野に参入し、様々な国際会議で発表しているということです。でも、先程指摘しましたように、アルゴリズムを発表しても、

計算力とデータ量が乏しい組織では、応用実践はできないという現実もあります。

■画像生成とディープラーニング

第3のディープラーニングとしてGAN (Generative Adversarial Networks, 敵対的生成ネットワーク)があります。これは2014年に提案された、画像生成用のディープラーニングです。具体的には、与えられた画像群から似通った(嘘の)画像を作る応用が考えられ、現在、フェイク画像とかフェイクムービーが問題になっていますけれど、ここにGANが使われています。GANを使って、お互いに面識の無いAさんとBさんが、にこやかに会話している動画を作ることができるのです。数年前は、本当の画像かフェイク画像か、簡単に見分けることはできました。しかしながら、現在、GANの性能は高くなり、例えば、前オバマ大統領とトランプ大統領がにこやかに話をするようなフェイク動画が、精巧に作成されており、本当ににこやかに話をしているのかと、一瞬勘違いするようなものも出てきました。このようなGANの使い方に対して、倫理的、プライバシー的に問題があると、各方面から指摘されています。でもGANの良い使い方もあり、セキュリティ対策にアタックするウィルスをやっつける強力なワクチンソフトをGANで作る研究もあります。AIの使い方は、人間側の責任ですので、今後、様々な観点から考えていくべきかと思っています。

■データサイエンス

現在、データサイエンスという言葉が広

くメディアで使われるようになりました。20年前はデータマイニングと言っていたのですが、知らない間にデータサイエンスに変わっていました。ここで問題になるのは、今までに説明してきました、機械学習とかディープラーニング、および統計的処理に占める時間は非常に小さいということです。最初に与えるデータ、ここが整理整頓されてなくて、クレンジングする必要があり、そういうところに非常に時間がかかっています。この図は、ディープラーニングの一番最高峰の国際会議でグーグルの研究者が発表した図ですけれども、この小さい真っ黒のところ、ここがディープラーニングと機械学習にかかるコストです。あとはデータを集めるとか、データが正しいかどうか検証するとか、特徴量を作るとか、データの前処理に関係したところで、これらの部分に高いコストがかかっていることが、この図から分かると思います。

■データは新しい石油？

データは新しい石油である、というキャッチフレーズが最近出てきました。しかし、石油は、不純物を含む原油を精油して製造されるものですから、生データは、石油ではなく、原油なのです。生データは不純物だらけなのです。その原油を精製して、クレンジングして、不純物を取り除いて、きれいな石油にすること、及び、足りないデータがたくさんあるので他の有用なデータと組み合わせるといふデータ前処理がポイントです。このデータ前処理に、あるベンチャー企業の発表では、データサイエンス全体工程で8割も占めると言っていました。

企業によって違う、問題によっても違いますが、まあ、データ前処理に6割から8割のコストがかかっているという現実があるのです。

結果の後処理、これは出てきた結果を実際有用かどうか、人間が読み解き、評価するところです。これも2割から3割のコストがかかっています。機械学習には、1割程度のコストしかかからないのは、なぜでしょうか。機械学習にはPython言語がよく使われますが、機械学習用ライブラリーがかなり充実しているので、ライブラリーレベルの組み合わせも考えれば、機械学習アルゴリズムが実装できるのです。我々の学生実験時でも、Pythonライブラリーの組み合わせを変えて、簡単に機械学習のアルゴリズムを作っています。どこかの高校か高専で、実験演習でディープラーニングを始めたと聞きました。そのように、機械学習は広く普及しており、あまりアルゴリズムを考える必要がありません。新しい機械学習アルゴリズムを考案する場合には長時間かかりますが、応用レベルでは、データの前処理、学習結果を評価する後処理の部分にコストがかかっているのです。

■データマイニングの課題

データマイニングの課題をまとめますと、データ拡大（追加）、及びデータクレンジングに高いコストがかかります。今、自分が持っているデータだけでは通常不十分で、どのデータと関係させるべきかを考察する必要があります。学習結果、マイニング結果、これを説明しようとしても、普通はなかなか説明できません。機械学習は、基本

は相関をとっているだけなので、あちらの変数の値が上がるとこちらも上がる、それは何故か、人間は相関では納得できなくて、因果関係、こういう原因と結果の関係があるからこうなるでしょう、と言われないと納得できないのです。その因果関係は通常、潜在変数となっており、隠れています。その部分をきちんと読み解く、ということで、マイニング結果、学習結果を解釈することも非常に大変です。

いろいろデータマイニングをやりましたが、担当者の意見は普通辛辣です。「ああ、こういう結果ですか」という感じで、それを受け入れようとはなかなかしてくれません。私は「主観と客観の対立」と呼んでいますけれど、ある意味でそれを受け入れると、自分のやってきた業務プロセスを否定されるように思われる節があります。連携するべきで、決してその仕事を奪うとかそういう話ではないのですが、業務担当者はけっこう冷ややかです。組織の壁、部門が違えば、もう連携できません。現場では、そういう面倒なこともあります。しかし、経営の世界でビジネススクールの先生と話をすると、もう PDCA サイクルは終わったと言われます。私の所属している管理工学科では、PDCA サイクルの話しがよく関連する学科なのですが、PDCA はもう終わったと言われます。OODA (ウーダ)、聞いたことのある方も多いと思いますが、まず見る、observe、何をすべきか方向付けて新しい事を考える、orient、決心して行動を起こす、decide and act という OODA の時代になってきたということで、ディープラーニングとか機械学習との関連性も大きくな

ってきているわけです。

ディープラーニング自身にも課題は多くあります。特徴量は中間層で自動生成されるというけれど、例えば、このように車種を判定するのに、車の型なんだから車の特徴量から判定しないとイケないのに、背景に人間が映っているから車種はこれだとか、背景に海が映っているから車種はこれだとか、背景画像を根拠にして車種を決めるといって、常識的にはあり得ない判断が学習されることがあります。このようなとんでもない誤りを防ぐには、知識を使って制約を与えて入力画像を精査する仕組みが必要になります。そうしないと、このように、バックに人がいたらマツダとか、海が見えたらトヨタとか、わけのわからない学習をすることがあるのです。ディープラーニングでは、「えっ？ これって本当」と思うような結果が学習されていることがあり、ディープラーニングの結果を読み解くことが重要で、先ほどのマイニングの結果の解釈と同じく、これは人間が担う重要な役割になります。

■機械学習の普及

機械学習の社会実装があらゆる分野で進んでいますので、その実践例を説明しましょう。例えば小売業で、これは 20 年前の機械学習の最初の実践例？として有名になったものです。ウォールマートが POS データを機械学習で分析し、缶ビールと一緒に買う珍しい商品がある、というのです。週末、既婚者、若い男性、缶ビールというように、時間帯、顧客属性、商品属性に制約をかけて相関ルールを学習させた結果、缶ビール

と併買する商品は紙オムツという結果が出たそうです。実際、これは作り話に近いという話もあるのですが、日本の本当の話として、相鉄ローゼンさんの実践例があります。この写真の商品陳列は、20世紀の常識からという、好ましくない陳列です。ここにペットボトルがあって、こちらにプチトマトがあります。生鮮食品とペットボトルでは賞味期限が違います。賞味期限の異なる商品は近くに配列しないというルールが、従来のスーパーマーケットの常識でした。ところが、POSデータを相関ルールにより購買分析すると、プチトマトを買う人はトマトジュースも一緒に買っている傾向があり、それなら賞味期限が違っても、商品を並べて陳列してもいいじゃないかということになりました。その結果、3割も売り上げが伸びたそうです。最初に、この報告を聞いた時、機械学習がマーケット分析に新しい常識を提供し始めたなあと思いました。

機械学習のその他の応用例としては、犯罪を予測するPredPolが有名です。SVMを使っているようです。警察が保有する犯罪データベースは、詳細レベルで多数のデータ項目が含まれ、例えば、空き巣を予想させるために、この犯罪データベースをSVMにかけると、翌日この赤い地域に空き巣が起ころうという予想が出てきます。そうすると、翌日、警察はここで空き巣を待ち構えていて、その空き巣がやって来て空き巣をやり始めた瞬間に「御用」となるわけです。ロサンゼルスでは、PredPolを適用し、48%犯罪率が下がったそうです。犯罪がおよそ半分になったのです。PredPolは、世界

で一番社会貢献しているAIと言えるかもしれません。

また、別の応用例として、食ロス問題に関連して、スシローの取り組みがあります。年間40億のPOSデータに機械学習をかけて、1分後の売れ筋ネタはマグロ、15分後の売れ筋ネタはエビとか客の食欲を予測し、ディスプレイに表示された予測に従って、すし職人さんがネタを握るわけです。すし職人さんは、予測は本当に当たるのか？と思っているかもしれませんね。でも、この予測通りに握った結果、廃棄寿司量が75%も削減されたのです。スシローのAIは、食ロス削減により地球環境問題に貢献しているといえますね。

■サッカーデータマイニング

組織による機械学習の実践例を紹介しましたが、私が関与した、機械学習・データマイニングの実践例について紹介します。

まず、サッカーデータマイニングです。10数年前に、Jリーグから提供されるデータを利用して、試合戦術を考えるプロジェクトに参加しました。ただ、データにはボールをキープしている選手の位置情報のみなので、無理のある話でしたが、スポーツデータマイニングの可能性を探ることが目標でしたので、参加しました。

ある選手がどの位置でボールをどのように処理したのか、パスかキックか、そのようなデータの集合です。選手の周辺にどのような相手選手がいるとか、いわば、試合の文脈に関する、試合戦術に関連するデータは含まれていないので、まず、可能なデータの追加を考えました。試合当日の天候

とか風のデータを追加しましたが、殆ど役に立ちません。試合戦術立案には役に立たない可能性が高いけれども、とりあえず、所与データから勝ち負けパターンのマイニングを試みましたが、学習結果の意味を読み解くことができません。そのため、ある J リーグチームにこの学習結果を持参したところ、そのコーチは「何だ、この勝ち負けパターンは!？」と学習パターンの意味を考え始めました。そのコーチは、ドイツの 2 部リーグで選手だった人で、試合戦術を論理的に考えられる非常に明晰な人で「う〜ん、これらのパターンが読み解けないのは、我々がゲームを展開する時に考えている、観点がないからだ」というので、「その観点は何なのですか」と聞いたら、「パスする方向とか、最終的にここにパスしたらどのぐらいの時間が必要か、これは 1 秒でいけるかとか 2 秒でいけるかとか、ここで必要時間とパスの方向を考えますね。それから、ボールタッチ数、何回タッチしてここまで到達するのかを考えています。」というような説明が始まりました。そして、コメントされた観点のなかで、所与データから導出できる特徴量を整理し、データを拡充しました。サッカーボールの位置データから、ゲーム戦術を考える上で意味のあるデータに一步近づいたと言えます。また、そのコーチが、パスやキックやファウルなどの一つ一つのアクションも、状況に依存して意味が異なるということで、アクションの意味付けも追加してくれました。

このように新しいデータ項目を追加して、決定木学習を実行したところ、165 個の勝ち負けパターンルールが学習できました。

そのうち 25 個程度のルールに、コーチと監督が関心を寄せてくれ「この勝ちパターンを実践すれば、我がチームは強くなるかもしれない」と言ってくれました。そして「選手たちに一度説明してみる」というので、すごくその結果を楽しみに待っていたのですが、「山口先生、申し訳ない。全部、否定されました」と言われてしまいました。「何故、駄目だったのですか?」と尋ねますと、例えば「選手が左方向に展開するのは勝ちパターン」というようなルールが学習されたとすると、選手達からは「でも、それは俺たちのプレースタイルじゃないです。今さら、プレースタイルを変えろといわれても無理です。」というような返事が返ってきたということです。選手の立場に立てば、ずっと練習してきたスタイルがあって、それを否定されるようなルールは受け入れられないという話です。学習結果に関わる人達の都合を考えたデータも追加して学習させないと、学習結果は受け入れられないということですね。

■肝硬変データマイニング

これは、ある大学病院のデータで、肝硬変のレベルは 1 から 4 まであって、レベル 4 となると致死レベルになるので、レベル 3 から 4 に変わるタイミングをデータマイニングするプロジェクトに参加しました。

過去 20 年間、1000 人に近い肝硬変データがありました。ただし「毎月検査に来なさい」と言っても来ない患者も多くいて、その部分は欠損値になるわけです。そのような欠損値が多く含まれているデータであり、すぐに機械学習を実行できる状態では

ありませんでした。欠損値を平均値に置き換える処理をすることもありますが、その間に急激に値が上がっているかもしれないし、その欠損値を埋める、あるいは、外す処置の可否の判断などが一番大変でした。機械学習とか統計の話ではないのです。欠損値をどういう考え方で補填するか、また、検査周期も患者によって違いますから、これをどのように統一するか、医学知識を通して、このようなデータ前処理を実行することが必要なのです。データ前処理の詳細については、かなり細かい話になるので飛ばしますが、データ前処理後、決定木学習により学習された、肝硬変レベル3から4への変化を読み取る、ルールの一つをお見せします。このルールでは、GPT（これは健康診断でもよく測定する値です）、ビリルビンとチモールの変化に注目しており、ビリルビンの値が高くて、チモールの値が減少して、GPTが減少に転じると、レベル4に移行して可能性があるというルールです。大学病院の先生からは、「GPTは変化しないのが常識なんだけど」とコメントされ、「正しいかどうかは不明ですが、ビリルビンとチモールがこういう変化をすれば、通常変化しないGPTが減少する」という驚きの知見だということです。ルールは500個程度学習されましたが、このルールだけ関心をもってもらい、あとは全く見向きもされませんでした。今回、専門家による学習結果の解釈では、500分の1しか有効性がなく、全体的には、かなり辛いところがありました。

■レジ無し店舗 AmazonGO

ディープラーニングの応用として、レジ無し店舗 AmazonGO がシアトルとサンフランシスコで運営されています。天井に数百台のセンサーをつけて、お客さんが棚から商品を取ったら、ネット上のカートがあって、今取った商品が課金されていきます。商品を戻すとこの仮想カートからこの品目も消えるのです。こういう動画があり、非常に面白い試みです。通常のコンビニと比べて最終的には1.5倍の収益を目指しているそうです。このレジ無し店舗は他社も追従し、世界で普及しつつあります。

■世界レベルの自動運転

車の自動運転は、今後、AIの最大の応用になる可能性があります。運転には、「認知」（人が歩いているか否か、信号がどう変わったかなどの認識）、「判断」（運転操作の計画）、「操作」（実際の運転操作）という3要素があります。

「認知」は、普通のレーダーと比べると検出精度が高いライダーを使います。ライダーは、高価で800万円程度しますが、アメリカのベンチャーを中心に開発競争が繰り広げられています。このライダーで人や物体を認識できたとすると、このような画像が出てきます。

そして、次に「判断」を実行するわけですが、運転時は、当然、交通ルールを厳守する必要があります。この部分は、ルールベースで処理します。ルールベースで処理した結果と「認知」で認識された結果を統合して「判断」を行うこととなります。何秒後に交差点前で停止したら良いとか、ハンドルを切るタイミングをいつにするか

とか、そういう定量的な判断を出してきます。この判断部分を手作業で調整することが大変なので、現在、ディープラーニングを適用して、自動的により良い判断に変えていく仕組みが研究されています。

最後に「操作」です。これは実際にブレーキを踏んだりするわけですが、これもスムーズにする必要があります。スウェーデンのボルボが、何とレーシングカーのプロのドライバーに運転させて、そこから機械学習させて、プロの運転操作を学習させようとしています。そこまでやる必要があるのかとも思うのですが、プロ並みの運転操作を機械学習させる企業も出てきているのです。この学習もディープラーニングです。自動運転の判断と操作をディープラーニングでいかにうまく上達させるかが、競争されているのです。

2018年12月、ウェイモ(Waymo)と呼ばれるグーグルの系列会社が、アリゾナ州で、自動運転有料タクシーサービスを開始しました。1600万kmも予備走行実験を実施しています。1600万kmという距離は、アメリカの全道路距離の3倍程度の距離です。25都市の公道で1600万kmの走行実験をした後に、自動運転タクシーの有料サービスを開始し、5マイル(約8km)走って、タクシー料金は、7ドル(770円)程度です。アリゾナ州なので人口は少なく、所詮、田舎だから実施できている感じもありますが、「2020年には、都市部で自動運転有料タクシーサービスを実施する」と宣言するライバル会社が、中国ベンチャー企業を含めて、数社出てきました。けれども都市部での自動運転は、かなり困難で、この計画をすぐ

にキャンセルする会社も出てきています。歩行者が多くなると自動運転は急に難しくなり、ここに大きな壁があります。まず、自動運転の実践は、田舎からというのが現状です。

■自動運転死亡事故と熾烈な競争

自動運転の大きな問題として、死亡事故もすでに起こっています。テスラとウーバーの2社、特にウーバーは歩行者をはねて死亡させて、ウーバーの自動運転実験は、危険度が高いという理由で、それ以降、中止になっています。

では、中国の自動運転実験の動画を見てみましょう。

<ビデオ再生>

中国の自動運転ベンチャー会社ロードスターのCEOは、アップルで自動運転エンジニアを務めた後、中国に帰国し、このロードスターという会社を立ち上げました。こういうケースが多くあり、アメリカは自動運転の技術を盗んでいると怒っているわけです。でも自動運転ソフトウェアエンジニアとしての経験を生かしているだけで、アメリカの自動運転の特許を侵害しているわけではありません。彼らは、エンジニアの経験をベースにして、新しい自動運転のソフトウェアを開発しているのです。ただし、中国の自動運転ベンチャーの競争は熾烈を極めており、ロードスターは、すぐに潰れるだろうとも噂されています。今、お見せした動画は、2019年1月、NHKがやっと取材できて、NHKスペシャルで放送されたも

のです。このベンチャーが、もう潰れるかもしれないと噂されるほど、競争が猛烈なのです。

■カリフォルニア州 DMV レポート

自動運転が世界レベルで競争される状況下で、自動運転の可否ではなく、自動運転の性能比較、特に、安全性の比較が求められる時代になってきました。カリフォルニア州 DMV(陸運局)では、自動運転の安全性ランキングをつけるために、2015 年以降、自動運転開始から人間の運転手に交代するまでの距離「自動運転解除（ディスエンゲージメント）距離」を公表しています。

2018 年の公表結果を見ると、一番安全なのはグーグル系列の会社 ウェイモで、約 18,000km 走行可能です。18,000km 自動走行後に初めて「運転を代わって下さい」と車が人に依頼するのです。先程の動画のロードスターはこのあたりの順位で、300km 位です。2 位にはアメリカの GM(General Motors) Cruise が入りました。GM がすごいというより、アメリカの自動運転ベンチャー、オートクルーズが優秀で、GM がここを買収した結果です。GM クルーズも、グーグルのアリゾナ州での有料自動運転タクシーサービス開始に対抗して「我々は、来年、大都市で自動運転を開始する」と宣言しましたが、すぐにその宣言を取りやめました。人口が多い都市部での自動運転は難しいのです。でも、このカリフォルニア州の自動運転走行テストでは 8,000km 走行できて 2 位になったのです。3 位と 4 位もアメリカのベンチャーで、3,000km とか 2,000km 走行できています。Pony.AI は中国

のベンチャーで 5 位に入っています。6 位に日本の日産が入り、順位的にはまあ良いのですが、走行距離は約 300km です。ウェイモの 18,000 km と比較すると、安全性で 60 倍もの差が付けられているわけです。こういう現状があって、自動運転については米中がリードしています。あと、アップル社は、自動運転特許数が世界一なので、自動運転技術はかなり優れていると見られています。アップル社の自動運転走行距離は 3km 程度です。実は、自動運転車の停止条件は各企業に一任され、同じ停止条件ではないのです。アップル社は、様々なデータを収集したために、かなり厳しい条件で停止させ、そのために走行距離が極端に短くなっていると言われています。「アップル社は、密かにすごい自動運転技術を開発しているらしい」とか、そういう噂などが飛び交うような状況が、現在の自動運転の世界です。

■今後の課題

社会実装が進む、機械学習、ディープラーニングの今後の課題について触れたいと思います。

お配りした、今年マカオで開催された人工知能国際会議（英語で IJCAI と書いてイチカイと呼んでいます）で、色々な発表を聞きましたが、ブラックボックス AI とホワイトボックス AI という言葉がよく使われていました。今流行っているディープラーニングは計算力とデータ量が必要ですが、学習結果は人を超えるケースが出てきて、すごいものがあります。医療診断で説明したように、大腸がんの発見では、ディープ

ラーニングの方が医師より発見率が高くなっているわけです。そういうことから、人間の眼より、ディープラーニングの眼が勝り、非常に高性能になってきた。でも、何故そう認識したのか、例えば、海が映っているからトヨタというような馬鹿げた判断をすることもあるので、人間がチェックする必要があるけれども、ディープラーニングはブラックボックスで、その判断理由を説明できないのです。何故トヨタの車と認識したのかと尋ねても、一切答えてくれないのです。ニューラルネットワークの詳細な計算過程を提示されても、人は分かりません。一方、ホワイトボックス AI はディープラーニングに比べると性能は低いのですが、例えば、決定木学習でしたら、例えばこの条件が何十以上で、この条件が何十未満だったら、このプラントは故障している、というように説明することができる、ホワイトボックス型 AI になれる機械学習です。このように、人間が点検して納得して機械学習結果を使って、知識ベース推論をするような AI を考えることができ、ディープラーニングに比べると性能は低いのですが、人間にとっては理解できて安心して使える AI になります。このような AI を XAI, explainable (説明可能な) という単語の ex の X をとって、explainable AI という意味で、XAI という研究分野が誕生し、2017 年から DARPA (国防高等研究計画局) 研究資金援助を開始しました。ディープラーニング中心の XAI, 決定木学習とか説明しやすい機械学習、知識ベース推論を利用した XAI など、様々なタイプの XAI の研究がアメリカ 11 大学で進められ、今後、こういう傾向

が加速化してくのかなと感じています。

最後に、皆さんにお配りした 2 ページの資料は、2019 年 8 月にマカオで開催された人工知能国際会議 IJCAI2019 の概要です。一般発表はディープラーニングが圧倒的に多いです。画像理解の畳み込み系ディープラーニングでは SeNet の改良版、言語処理系ディープラーニングでは BERT の改良版の発表が、多くの中国人により発表されていました。一方、招待講演者はアメリカの先生が多かったです。ディープラーニングは必要だけでも、記号処理と連携させてホワイトボックスにしていく必要がある、推論システムの方をメインにして機械学習は一つのサブモジュールみたいにするべきという発表もありました。

ビッグデータを使った機械学習、ディープラーニングに基づくデータ学習型 AI が、現在の第 3 次 AI ブームの中心ですが、今後は、人と協働作業ができる説明可能な AI, データ型 AI と連携可能な知識型 AI, こういう AI が新しい方向になっていくのかなと、IJCAI2019 に参加して感じた次第です。それでは、以上で講演を終わります。

(文責：編集委員会)