

(2017 年 1 月 18 日ご講演)

AI（人工知能）の現状と展望

株式会社 KDDI 総合研究所
リサーチフェロー 小林雅一委員

1. 現代 AI の概要

私は、人工知能の歴史をルール・ベースの AI、統計・確率的な AI、脳科学の成果に基づく AI、この 3 つのフェーズに分けて考えている。厳密にきっちりと分けているのではなく、この 3 つがない交ぜになっている。ただ、その進化としてはこういう方向性で来たということである。

1960 年代～80 年代の当初のものは、いわゆるエキスパートシステムや、人間社会に蓄積されたいろいろな知的産物をコンピューターが分かるようにルール化して、プログラム言語で移植していったような形で、柔軟性に欠けていた。例えば、人間は文法どおりに言葉を話しているわけではない。機械翻訳をするにしても、現実世界に柔軟に対応できなかったということであまりうまくいかなかった。ただ、いまだにこういったものは、siri にしても Watson にしても、それが改良されて当然使われているわけである。基本的にはこういった形で、一時期 AI の冬と言われているような時期を経て、その後、一般的にはベイズ定理を使った統計的な手法が 1990 年代後半から主流になるが、いまだにこれが結構使われている状況で、いわゆるベイジアン・ネットや隠れマルコフモデルなど、そういったものをベースに作られているのが、今の基本の主流となっている人工知能である。

今後の動向は、ニューラルネットが主体になっている。この業界で今非常に流行語になっている Deep Learning といったものは、かなり古くから研究開発は続けられてきたが、今世紀に入って、恐らく 2004 年ぐらいが分岐点だったと言われているが、そのあたりでいろいろなブレークスルーがあり、今ブームを巻き起こしているということである。いろいろ統計的な手法を使っていることは基本的に同じであるが、昔と違うのは、本格的に脳科学の成果が人工知能に導入されてきたというのが最近の非常に大きな傾向である。

いずれどのような形になるにせよベースにあるのは、いわゆるビッグデータと言われているもので、現実世界に発生する大量のデータを処理するために今の主流の人工知能があるし、逆に、それを処理することによってどんどん性能が上がっていくというような持ちつ持たれつの関係にあるわけである。

こうした AI を構成する要素技術を次に示す。これだけではないが、いろいろなものがある。音声認識、画像認識、自然言語処理、等々。自然言語処理と音声認識はどこが違うのかと言われるかもしれないが、音声認識はただ単に音声を認識して、人間が使う場合には

当然言葉というよりも音声だと、文字にするところまで。自然言語処理は、文字になったものを意味ある言葉として認識するということで、そこにはだいぶ大きな谷間がある。あるいは機械学習というのは、ビッグデータの中のある種の法則性とか規則性、パターンといったものを抽出するための技術で、今これが非常にブームになっている。確率的推論というのは、物事と物事の因果関係を確率的に記述する。よく例に出されるのが、ある朝窓のカーテンを開けたら芝生がぬれていたと、理由は確率 70%で少し前に雨が降った、確率 30%で誰かがスプリンクラーで水をまいたと、これがいわゆるベイズネットの典型的なものであるが、こういったものが要素技術としてあるわけである。

これ以外にもたくさんあるが、こういったものを組み合わせることによって既にいろいろなものが製品として出てきている。例えばテレビに向かってこういう番組を見たいと言葉で言えば、音声認識とか自然言語処理といったものを使ってテレビがそれをある程度理解して、見たいテレビ番組をポンと出すとか、掃除ロボットも一種の AI かもしれないし、介護用ロボットなどはこれからであるが、いろいろな物が造られているということである。

家電のほかに、自動車、軍事、医療、介護、金融・投資、法律、農業、教育など、非常に多岐にわたる産業に应用が期待されており、非常に AI が注目されているということである。これをどう見るかであるが、今のところマスメディアの捉え方としては、非常にネガティブに捉える傾向が強く、例えばこういったことで AI やロボットが出てくると、人間の知的労働、作業までそういった AI やロボットに仕事を取られてしまうのではないかと。その問い掛けはもっと前向きに捉えることもできて、日本をはじめ欧米もそうであるが、先進国は産業が非常に成熟化している。成熟化しているということは、製品が見飽きたものになってしまったということであり、私たちの子供のころからテレビも冷蔵庫も車もあるし、今の若い人に見れば、車に乗りたがらないというものも、もう見飽きた製品になっているといえる。だが、そこに人工知能が搭載されることによって、今まで見飽きた製品が全く新しい物に生まれ変わり、産業が再び活性化されるということ、そういう前向きな見方のほうがよいのではないかと思う。

2. 自動運転車

中でも特に注目されているのが自動運転車だろう。それはいろいろな理由があるが、特に日本や欧州諸国の場合、自動車が基幹産業であることと、産業の裾野が非常に広く、下請け孫請けなどがいて、多くの雇用を創出する産業でもある。そこでここが人工知能によって全く新しいものに生まれ変わるということは、産業的なインパクトが非常に大きいわけである。だからここが非常に注目されているということである。

スポーツカー Audi の A7 で、ホッケンハイムのレース場を自動運転で周回するというデモをしているが、ニューラルネットの Deep Learning といったものを搭載して、要するになるべく精度の高い形で外界を認識しながら走る。あと機械学習が当然入っていると思うが、最短コースが一番速いわけではなく、少し大回りしたほうが速かったりすることがあ

るではないか。そういうものを自分で学習する。

自動運転車は各社が 2020 年ぐらいを目途に製品化したいと言っているが、これは少し難しいかと思う。実際に 2020 年までにとなると、それほど簡単ではない。GM(ゼネラル・モーターズ)は 2019 年にはハンドルの無い自動運転車をリリースしたいとしているが、米国の規制当局が許可するかどうかは未だ分からない。ただ、一説によると、いろいろなやり方がある、例えば Google などが考えているのは、最初は当然事故があるので、それを織り込んで、例えば本当に小さな自動運転車で、バンパーの部分に柔らかいクッションのような物を付けて、街乗り限定し、速度は恐らく最高でも時速 30 km ぐらいにしておけば、仮に人にぶつかってもそれほどひどいことにはならないだろうということ、要するに自転車ではないが、タクシーの代わりに使わせるといったところから恐らく入っていくのではないかとされている。主導権争いは、また言わずもがなである。

いろいろなセンサーが入っているということで、波長の違う、反射してくる物体の性質が違って、例えば人間のような有機物に反射する波長や、金属、プラスチックに反射する波長など、いろいろなセンサーを使うことによって総合的に外界を把握するというのである。ここには、いわゆる初期のルール・ベースの AI から先ほどのベイズ型の非常に統計的な形、確率的な形の AI から Deep Learning など、そういったニューラルネットもどんどん入ってくる方向によって、いろいろな人工知能の最先端を集めている。

行動計画を立てるというのは、昔のやり方に近いと思うが、今日は自分の周りにいる歩行者や障害物をどのように自動運転で把握しているかを簡単に紹介していく。そうすると、いわゆるベイズ定理を使った現在主流の人工知能がどういったものかお分かりいただけると思う。

ベイズ定理というのは、私が高校生のころの教科書では条件付き確率とか言って教えていたような気がするが、そのことである。主観確率とか、そういう考え方に基づいているということである。私たちが中学生のころに習う確率はいわゆる客観確率である。さいころを 6 万回振ったら 1 から 6 の目の出る回数は約 1 万回ずつになるというのが客観確率の考え方であるが、それはあまり役に立たないことが多い。どうしてかというと、現実の状況として 6 万回もさいころを振っている時間はないし、大変である。だから、トーマス・ベイズが考えたのは、もっと主観的な確率で構わない、人間というのはもっといいかげんに確率を考えていて、例えば私はテニスが好きであるが、初対面の人と会ってテニスをするとき、チラッと相手を見て、少し顔色が悪そうだとか、あるいはお腹が出ていたりしてあまり運動をやっていないのではないかとすると、私の勝つ確率は大体 70% ぐらいだと。大した根拠はない、これが主観確率である。それがここで言う事前確率である。

ただ、それだけでは心もとない。どうするかというと、試合を始める前にポンポンと相手と何回か打ち合ってみる。それはウォーミングアップであるが、ある種のテストでもあるわけである。相手の実力をその間に推し量る。それがある種の測定とか、ここで言う実験とか観測に当たるわけである。その情報を加味することによって事後確率、若干精度の

高い確率が得られるということである。つまり、ラリーをする前は 70%ぐらい勝つと思っていたのが、ポンポンとラリーをしているうちに、どうも相手は意外に強そうだという感触を得れば、自分が勝つ確率が 65%に修正されるわけである。それが事後確率である。

ベイズ定理というのは、結果から原因を推定するための定理とよく言われるのだが、同じことである。私が今言ったことと、結果から原因を推定するということは実は同じことを言っていて、結果というのは、その場合、実験の結果ということを行っているわけである。そこから事後確率を出す。それは実験をすることによって相手の実力がより正しく分かることができるようになる。まさしく実験の結果から本当の原因、つまり相手の力量というものが分かる、それが事後確率である。

自動運転車はまさにこれを使っていて、まず当然常識的にいそうな場所というのは、自動運転車から見た、例えば周りにいる車、歩行者、あるいは何かの障害物、移動体である。ほかの車の場所だと思ったら、それは当然正規分布で表されるわけである。

たくさんのセンサーがあっても、1回計測しただけでは当然誤差が大きいわけである。本来 3次元であるが、1次元の世界を考えるとしたら、このあたりに歩行者がいそうだなというのはセンサーで測ってみると分かる。確率が一番高い。当然正規分布。ただし、誤差が大きいわけである。

最初の状態では大体歩行者は 1次元の世界で、まだこのあたりにいる確率が結構高いと分かっても、誤差が非常に大きいので危ない。だからもう 1回測定して先ほどのベイズ定理を適用して事後確率を得ることによって、その誤差の幅を小さくするわけである。精度の高い確率が得られる。それを 1秒間に何万回やっているのか分からないが、それこそ何万回、何十万回とやることによって急速に誤差が収束するわけである。そうすると、最終的にこの 1次元の世界で歩行者がここにいることがわかるわけである。自動運転車は、まさしくここに人がいるのだから、ここだけ迂回していけばよいわけである。これがベイズ定理を使ったカルマン・フィルターの原理である。ロボット工学でよく使われる原理であるが、こういった形でやっている。これはまさしく今主流の統計的な手法に従う人工知能である。

これだけではない。自動運転車だけではなく、例えば少し前までは Google 翻訳も基本的には恐らくこのベイズ定理を使ったやり方でやっているが、いろいろな分野でこういったことが使われている。これを人工知能と呼ぶかどうかは、その考え次第であるが、使い次第で、人間の考え方や行動に近いようなものを、こういった統計的な手法で実現できるということである。

ただし、これには問題が含まれている。正規分布を使っているところが非常に危なくて、自動運転や人工知能だけの世界の話ではなく、一般的にこれが問題視されるのは金融の世界である。定期的にいわれる金融危機、世界的な金融恐慌というものが起きるのは、原因がこれにあると言われている。正規分布というのは基本的に数式として使いやすいので、例えば金融工学といったところに使われているのだが、本当にそれが正しいのかというと、

どうもそうではないらしい。そういう研究をしている数学者やエコノミストに言わせると、これは一般的に言われていることであるが、正規分布と現実世界とはずれていて、現実世界はどちらかというといわゆるテールの部分が少し確率が高い。要するにファットテールや、ブラックスワンと言われているが、正規分布上は確率はゼロに近い部分だが、現実世界は実は正規分布からずれていて、この端のほうの確率が結構高い。

どういうことかということ、例えば一番直近の大きな世界的な金融危機というと、2007年ぐらいのリーマンショックである。あの引き金と言われているのがアメリカのサブプライムローン、低所得者向けの住宅ローンであるが、正規分布上は前提としてあれぐらいの規模の貸し倒れが起きるといえるのは、100万年に一度ぐらいのほとんどあり得ない事象で、まさしく正規分布上は事実上ゼロである。ところが、起きてしまったわけである。その10年前、1997年、98年ぐらいのアジアの通貨危機で、結果として当時の大きなヘッジファンドのLTCMが破綻したが、その引き金となったのがロシアのデフォルトである。ロシアのデフォルトも、正規分布上、金融工学上は起きない、100万年に一度ぐらいのことが実は起きてしまったわけである。そのさらに10年前が1987年のブラックマンデーである。つまり、理論上は起き得ないことが実際の世界では、それほど高い確率ではないのだが、100万年に一度ではなく10年に一度ぐらいで起きているということである。だから怖いのは、これと同じ原理を自動運転車が導入しているということである。

一番分かりやすい例でどういうことかということ、自動運転車から見たほかの車の移動であるが、邪魔な車が $T=t_0$ 、 $T=t_1$ のときにいる場所から、これは当然等速直線運動だろうと、したがって $T=t_4$ のときにはここにいるだろうと自動運転車は予想するわけである。正規分布上は当然そこである。だから問題ない。要するに自動運転車は、危ないほかの車がいるところには絶対に行かない。問題は、微妙な判断を迫られるケースである。つまり等速直線運動をしていた車が、あるときギアを急にバックに入れて逆走する確率は、正規分布上はほぼゼロに近いとみているわけである。ところが、実際は結構起きる。カルマン・フィルターは正規分布的な予想をベースにしているので、安心して自動運転車はそこへ行く。そうすると、事故が起きてしまうわけである。

これはただ単に理論の話なわけではなく、実際にこういうことが起きている。昨年2016年5月ぐらいにTeslaのAutopilotでの事故があった。彼らは自動運転とは言っていないのだが、Autopilotという名前自体が自動運転である。彼らはドライバー支援システムで、自動運転ではないから、あくまでも事故を起こしたときの責任はドライバーというかユーザーにあると言っているのだが、彼らは自動運転として売り込んでいた。非常に無責任である。関係者がいたら申し訳ないが、多分ご存じだと思うが、AutopilotはTeslaが2015年の秋ぐらいにリリースした機能で、無線インターネット経由でTeslaのModel Sという車にダウンロードできる。その機能をダウンロードすると、普通の市街地だと信号などがあり、ほかの車もいろいろな運転をしていて難しいので、原則的には使用環境をハイウェイに限定してほしいということが使われた。手を離しても運転できるのだが、Teslaが推奨し

ているルールとしては手放し運転はせず、いざというときには、ステアリングに手を掛け、アクセルやブレーキングを使えと。だが実際のユーザーは、普通の道で使ったり、事実上の自動運転として手放し運転をしたり、車内でビデオゲームで遊んだり、そういう危ない使い方をしていたわけである。この事故は、フロリダ州の北を走っている州間ハイウェイを時速 100 キロぐらいで Autopilot 走行していた車が、対向車線から左折してきた大型トレーラーの側面に衝突したものである。

3. Deep Neural Network (DNN: Deep Learning)

このような事故を回避するためにどうするかというのは、いろいろなやり方があり、単純に個別型のやり方だけではなく、先端的なニューラルネットも入れていこうという方向にあるわけである。だから、特に自動運転車のせいで Deep Learning などが注目されているわけではないが、そういったこともありより精度の高い人工知能が求められ、Deep Learning やニューラルネットが再び今改めて注目を浴びている。

1950 年代から研究開発が始まったが、その理論自体は 1940 何年に神経生理学者や数学者などから始まったと思う。人間の頭脳には 1,000 億個ぐらいのニューロン（神経細胞）があるということで、その動きをシグモイド関数など、今は別の数式が使われているが、要するに非常に初歩的な数学の数式を使って表現し、それを組み合わせて人間の脳のようなものを人工的に造っていこうということで始まった。

初期のものは二層構造である。Perceptron と言われているもので、これは今も使われている。統計の専門家はロジステック回帰分析と呼んでいるが、単純に線形分離と言って、本当に一番単純な分離の仕方である。1本の直線を引いて、上にあるものと下にあるものにグループを分けるということで、そういう考えは結構使い勝手があっただけに使われているものであるが、作られたのは 1950 年代である。初期のものはハードウェアだった。今もウェブで検索すると写真が出てくるが、とんでもない大型のコンピューターだったが、このような簡単なことでも、当時はあのようなすごい設備が必要だったのかなという感じである。

原理的な問題が指摘されて、二層構造だと排他的論理和が理解できない。排他的論理和というのは、ベン図で書くと、P または Q なのだが、P かつ Q の部分だけは抜くと、小学生でも理解できるようなことが理解できないことが分かって、いつか 1950 年代に開発ブームが巻き起こったが、簡単なこともできないというので下火になってしまった。だが、それからいろいろ辛抱強く研究する人たちがいて、層の数を増していくと、どんどん難しいことも理解できる。三層構造にすると排他的論理和というのは処理できるようになる。類推で、たくさん層の数を増やしていけばどんどん難しいことができるようになるということで、これがいわゆる今のディープ（多層）・ニューラルネット、つまりディープラーニングの始まりである。ただし、本当の意味で脳をまねしているのかというと、そうではなく、ニューロンの動きをシグモイド関数でシミュレーションするというのは確かにある意

味ではそうであるが、基本的に 99% ぐらいは数学的な処理というか、数式の産物である。

そういうわけではないのだが、多層化すればするほど数学的な処理に非常に時間が掛かって、結局データ処理に時間が掛かるということであり、また汎用性に乏しいと言われていた。音声認識用に開発されたニューラルネットは、一例としては画像認識には使えない。だから、非常に金を掛けて何人もプログラマーを導入して作っても 1 個の目的にしか使えない、あるいは遅いといったことでいったんブームが盛り上がり、また下火になり、またやるのではないかとということでまた盛り上がり、また下火になってという繰り返りで、駄目かなと言われた時期が結構長かった。21 世紀に入っているいろいろなブレイクスルーが起きてきた。例えば、どこかの大学院生が最初に言ったらしいが、専用の画像処理用のプロセッサをたくさん使うと非常に処理が速くなる。そういったことによって速度の問題が、要するにハードウェアの進化によって解決されてきた。

それから、汎用性の問題だけではないが、このころから、21 世紀に入ってようやく本当の意味でのニューラルネット、脳科学の成果が本格的に導入されてきた。それ以前は単なる数学の産物に過ぎなかった。一例としては、ニューラルリワイヤリングという実験がある。1990 年代に世界中の脳科学者がいろいろ奇妙な実験をやった。残酷であるが、例えばフェレットというイタチのような動物の頭蓋骨を割って脳細胞を出す。そして視覚情報を処理する領域につながっている神経のラインを切断して、無理やり聴覚野につなぎ直した。そうすると、そのフェレットはかわいそうだが当然目が見えなくなってしまう。ところが何カ月かすると、またぼんやりとはあるのだが、どうやら目が見えるようになったという、これは有名な実験らしい。ここから脳のいわゆる感覚野というか、物を見たり聞いたりする領域に通底するある種のアルゴリズムというかセオリーが発見されたということである。脳には意外に汎用性があり、無理やり視覚情報を聴覚野に入れてやると、聴覚野が視覚情報を処理できるようになる。それを数式で表現できるようになった。Sparse Coding などと言われているものである。

こういった形のものを AI の研究者が聞き付け、そういう便利なものがあるのだったらと人工知能に導入した結果、人工知能の画像認識の能力が飛躍的にアップした。例えばスタンフォード大学主催だったか、世界的な画像認識のコンテストを年に数回やっているらしいが、そういうコンテストでは画像認識用のソフトウェアに対していろいろな課題を出すわけである。何十万枚もの写真を見せて、実際は数値を入力するだけであるが、例えばチーターとレパードなど人間が見ても分かりにくい画像を、この Deep Neural Net は人間以上の精度で正確に把握できるようになる。つまり画像認識の精度が飛躍的に上がったということが、たしか 2006 年ぐらいに証明された。

なおかつ、先ほどの統一的な情報処理の仕組み、汎用性があるものの原理を取り入れているので、恐らく画像認識でそれだけ精度が上がったということは、音声認識にも使えるのではないかとということで、実際にやってみたら非常に音声認識の精度が上がった。例えば Google ブレインというのは Google がやっていた Deep Neural Net の変形であるが、そ

ういったところの成果をスマホの音声検索に導入したところ、音声認識の精度が非常に上がったわけである。だから、私たちが最近普通に iPhone や Android から音声検索ですると非常に精度が上がったが、まさしくこういった原理が使われているわけである。それによっていわゆる画像認識や音声認識といったパターン認識が飛躍的な進化を遂げたということである。これをこれからもっといろいろな分野に応用していこうという状況に今あるわけである。

4. DNN : 産業への応用と諸問題

一例としてはロボットに使っていこうということで、2013年ぐらいにロボットのコンテストがアメリカで開かれた。いわゆるヒト型ロボットだったが、あまりうまくいかなかった。2013年が予選会で、2年後の2015年ぐらいに本選があり、最終的には韓国のチームが優勝したのだが、しょっちゅう転んだり、止まったまま動かなかったりと、あまり大したことがなかった。恐らく外界を認識する能力に非常に限界があったことが、その大きな理由である。動こうと思っても、ビデオカメラなどで外界を撮影して正確に外界を認識しないことには動けないので、その部分で時間を取られてしまった。そこで今は、パターン認識にすぐれた人工知能をどんどん導入していこうという方向で、いろいろな大学の研究などを行っているわけである。

今のところ AI は、ニューラルネットにしてもソフトウェアとして実装されることが多いが、1チップ化してハードウェア的にやっっていこうといった方向で、これは IBM 等も非常に力を入れているところである。将来的には家事を助けるロボットや、いろいろな物を実現していきたいと思う。

それから、自然言語処理であるが、昨年2016年11月に Google が Google 翻訳という機械翻訳のサービスをリニューアルした。昔のやり方、恐らくベイズ型のものからニューラルネット Deep Learning を使った形にリニューアルして非常に評判が良い。昔はウェブ翻訳はほとんど使い物にならないと言われたが、非常に評判が良い。実際に使ってみれば、大体今の自然言語処理のレベルがどういったものか分かると思う。やっていることはいわゆる回帰分析である。例えば右側にドイツ語の文章を入れて、左側にそれを人間が翻訳した日本語の文章を入れて、ただ単に右と左を比較して、このあたりにある単語は、日本語だとこれになっているからこれなんだろうと。それを何万回もやることによって、だんだん確率を改良していくことによって最終的に翻訳文ができる。つまり意味は全く理解していない。単に統計的な処理でやっついて、Deep Neural Net もその点は同じである。だから、構造が近い言語、例えばドイツ語と英語、英語とフランス語だったらうまくいくが、多分日本語とドイツ語、あるいは韓国語とドイツ語ではうまくいかないと思う。

また、プロモーションビデオなどでは、家庭に浸透して家族と仲良くやっているロボットが紹介されたりするが、実際はまだ会話ができるような状況にはなくて、限界があるわけである。Geoff Hinton という有名な研究者がアメリカの大学の講演で、今のニューラル

ネットの自然言語処理の能力をデモした動画がある。**What is the meaning of life?** と、人工知能に聞く。機械学習をして大量の文献を読み込んで、ある種できる範囲で言語能力を得た人工知能ニューラルネットに対し、人生の意味は何かと聞くと、スクリーンにAIが返してくる答えがパラパラッと表示されていく。それが一見哲学的であるが、何か親愛があったり、文法的にもまあまあ正しいと。だが、きちんと見ると明らかにこれはナンセンスなことを言っている。要するにまだとても会話に使えるようなレベルにはなっていない。つまり、ニューラルネット **Deep Learning** は、パターン認識まではうまくいったが、その先の人間のように言葉を理解して会話するところになると、まだまだ非常に限界がある。