

# 経営講演会

## 講演録

『加速する人工知能の進化と産業・社会への影響』

— ディープラーニングが切り開く新時代 —

(平成28年11月7日 講演)

講師 東京大学大学院 工学系研究科 特任准教授

松尾 豊 氏



りそな中小企業振興財団



講師

東京大学大学院 工学系研究科 特任准教授  
松尾 豊 氏

◆プロフィールご紹介

主な経歴： 2002年 東京大学大学院 工学系研究科電子情報工学 博士課程修了、博士（工学）  
2002年 独立行政法人 産業技術総合研究所 研究員  
2005年 スタンフォード大学 CSLI（言語情報研究センター） 客員研究者  
2007年 東京大学大学院工学系研究科 総合研究機構（若手育成プログラム  
スーパー准教授）／知の構造化センター／技術経営戦略学専攻 准教授  
2014年 東京大学大学院工学系研究科 技術経営戦略学専攻 消費インテリ  
ジェンス寄付講座 共同代表・特任准教授  
2015年 産業技術総合研究所 人工知能研究センター 企画チーム長（兼任）

この講演録は、平成28年11月7日に開催された、当財団主催の経営講演会を収録・編集したものです。なお、財団ホームページにも掲載しております。<http://www.resona-fdn.or.jp>

今年の3月に、**ディープマインド社**というグーグルが買収した人工知能の会社がつくった**アルファ碁**が、韓国トッププロを相手に4勝1敗で圧勝しました。これは人工知能関係者にとっても非常に衝撃的でした。

実は、囲碁は非常に難しいゲームで、将棋の場合、人工知能はここ2~3年でようやくトッププロに勝てるようになってきたので、囲碁でトッププロに勝つのは2025年ぐらいかと思われていました。最大の勝因は、**ディープラーニング**という技術を使ったことです。

**ディープラーニング**というのは画像認識に非常に強い技術で、このアルファ碁の場合は盤面を画像として捉えて、画像認識の技術によって盤面の形勢を判断させた。それに加えて、自分同士の対戦を3,000万局ぐらいやって、強くなったということです。このディープラーニングについて、お話をしていきたいと思います。

## 1. 人工知能ブーム

いまの人工知能ブームは、歴史的には3回目のブームです。人工知能の分野は1956年にできたと言われているので、今年でちょうど60年目です。この間、ブームと冬の時代を繰り返してきました。

人工知能は非常にブームになりやすくて、人間のように賢くなるだろうという想像をいつの時代も人々は持ってしまいます。第一次のAIブーム(1950~1960年代)のときも、世の中では人工知能が人間よりも賢くなるという期待がありましたが、うまくいかずに失望に変わりました。第二次のAIブームも日本ではエキスパートシステムというのがかなり使われて、当時の通産省は「第5世代コンピュータプロジェクト」でお金を注ぎ込みましたが、うまくいかずに失望に変わってしまいました。

今回もかなり期待感が高いのですが基本的には60年前からやっている技術が少しずつ良くなっていまに至っているのです、いままでできなかったことが急にできるようになるわけがない。

キーワードとしてよく出てくるのが**IBMのワトソン**で、クイズの世界チャンピオンに勝って有名になりました。あるいは、自動運転。ソフトバンクの**Pepper**。将棋・囲碁。**iPhone**に入っている**Siri**。でも、昔からある技術なので、急激に変わってはいません。

例えば、**IBMのワトソン**は銀行のコールセンター業務や医療の診断に応用されましたが、人工知能の医療応用は、実は1970年代から行われています。1970年代初頭、スタンフォードにマイシンという非常に有名なプロジェクトがありまして、「血液検査結果がこうだとこの疾患の可能性が何パーセントだ」というルールを500個ぐらい書くと、大体当たるようになり、当時でも研修医の診断精度を超えていたと言われています。

では、なぜ未だに使われていないかということ、人間の医者には、血液検査の結果以外にも患者の顔色、病歴、職業等、いろいろなことを総合的に判断しているわけです。この「**総合的に判断する**」ということが、コンピュータは非常に苦手です。ですから、ある程度のところまではすぐに行きますが、そこから精度を上げることは非常に難しいのです。ですから、知識を入れるタイプの人工知能、ワトソンにも同じ特徴があります。

いまは人工知能全体に対しての期待感が過剰になっていますが、ディープラーニングだけは別格です。ディープラーニングに関しては、何十年もできなかったことがこの2~3年で次々とできるようになっています。ここは、期待してもし過ぎることがないぐらい潜在的な可能性が高い領域ですから、投資するのであればここしかないと思っています。

## 2. ディープラーニングでできること

ディープラーニングで一体何ができるか。非常に簡単に言うと、まず**認識**ができるようになる。次に、**運動の習熟**ができるようになる。そして、**言語の意味理解**ができるようになる。この3点です。

**認識**というのは、画像認識です。コンピュータは画像認識が非常に苦手でしたが、これができるようになってきました。**運動の習熟**というのは、ロボットや機械が練習して上達することができるということです。**言語の意味理解**は、言葉の意味を本当にコンピュータがわかるということです。これを、順番に説明していきたいと思います。

**画像認識**はコンピュータにとってすごく難しいタスクです。人がネコ、イヌ、オオカミの3枚の写真を見ると一目でわかりますが、これがコンピュータにはすごく難しい。

なぜかという、例えば、目が丸ければネコと判断しよう。目が細長い場合に、耳が垂れているとイヌで、耳が尖っているとオオカミと判断しよう。このようなルールを作ると、うまく判断できるように思えますが、例外が必ず発生します。オオカミっぽさを定義することはすごく難しい。

目が丸いとか、オオカミっぽいというものを**特徴量**と言い、特徴量を人間が定義している間は画像認識の精度は一向に上がりませんでした。これが、機械学習における特徴量の設計が難しいということです。

人工知能には、それ以外にも**フレーム問題**とか、**シンボルグラウンディング問題**という、昔からの難問があるから、人間のような知能はなかなか実現しないのだと言われていました。しかし、根本的な問題は1つです。それは、従来の人工知能は人間がモデルを立てており、その行為自体は自動化されなかったことです。

これを解決しつつあるのが**ディープラーニング**で、僕は「50年来のブレークスルー」と言っています。いままでの人工知能は人間がモデル化した後の自動化のことを言っていますが、**ディープラーニング**はモデル化を一部自動化しているので、次元が違うわけです。

そのやり方は難しいので飛ばしますが、有名なのが「グーグルの猫」という研究で、たくさんのインターネット画像をニューラルネットワークというタイプの人工知能に見せると、入力に近い側のレイヤーでは線とか、点のようなものが学習され、もっと右のほう、上のほうに行くとヒトの顔っぽいものやネコの顔っぽいものが自動的に学習されるやり方が見つかりました。つまり、「ネコらしさ」が、画像をたくさん入れるだけで学習されるということです。これは、人間の脳の視覚野の仕組みと非常に近い仕組みです。

このグーグルのネコの研究がされた2012年に、ディープラーニングはアルファ碁と同じ

ような衝撃を画像認識の分野で起こしました。ヒョウ、コンテナ船、プラネタリウム、コアラなどの写真を見て、それが何かを当てる問題を数万個解くことで、画像認識のエラー率を判定します。

2012年直前の2年間は、エラー率が大体26~27%でしたが、2012年にはディープラーニングのチームが16%を出し、圧倒的に精度が高くなりました。しかも、他のチームは、特徴量の設計を一所懸命改良して少しずつ精度を上げていましたが、ディープラーニングのチームはそれを一切やらずに勝ったので、衝撃的でした。

2012年以降、すごい勢いで画像認識のエラー率が下がってきていて、人間は5.1%間違えますが、最新の手法だと3.6%で、今月は3.1%が出ました。つまり、2015年2月に、コンピュータが画像認識で人間の精度を超えてしまったのです。これは歴史の教科書に載るのではないかと思うぐらい大きな出来事です。

### 運動の習熟

次に起こっている変化が、**運動の習熟**です。つまり、練習して上達するということです。人間は、例えばゴルフボールを打っているとだんだん上手くなってきます。なぜかというところ、たまたま上手に打てたらそのやり方を繰り返すわけですが。報酬が与えられると、その前にやった行動を強化して学習する仕組みなので、**強化学習**と言われます。

強化学習の仕組み自体は昔から知られていて人工知能でも随分研究されていましたが、例えば、ゴルフでも、「こういう状況ではこういう打ち方がいい」というふうに、「状況」と「行動」をセットにして学習させていく必要があります。

ところが、いままでの人工知能のやり方では、「こういう状況」の記述に、人間が定義した特徴量を使っていました。しかし、ディープラーニングと組み合わせる方法では、ディープラーニングで出てきた特徴量を使って状況を定義します。そうすると、想定外の状況を上手に認識できるようになります。状況の定義をディープラーニングにやらせるだけで、非常に大きな変化が起こります。

幾つか動画をお見せいたします。最初は、アルファ碁をつくった**ディープマインド社**が2013年ぐらいに研究していた、ブロック崩しを学習する人工知能です。強化学習なので、最初は下手な状態でだんだん上達してきます。スコアを報酬にしてあるので、スコアが上がるとその前にやった行動を強化して上達します。このぐらいはいままでのAIでもできました。

これがすごいのは、いままでの人工知能と違い、ボールだとか、自分の動かすバーという定義を一切せずに画像を入れているだけだという点です。そうすると、「丸っこいもの」とか「バーっぽいもの」とか、「その位置関係がこうであるときに点が入りやすい」というのを学んでいき、そのうちに端に通路をつくって点を稼ぎ始めます。これは画像から特徴量を出して、左端や右端に通路ができた状態がいい状態だということに気づいているわけです。これは、いままでの人工知能のやり方ではできなかった。

更に、同じプログラムを使って、全然違うゲームを学習させることができます。これも

驚くべきことで、いままでは、インベーダーゲームを学習させようとする、「インベーダー」とか、「ミサイル」を人間が定義しないといけなかったのですが、これは画像を入れてスコアを報酬としているだけなので、インベーダーゲームも学習させることができます。

これは昔のアタリ社のゲームですが、60種類ぐらいあるうちの半分の30種類以上の、反射神経、運動神経だけでいいものは、人間のハイスコアよりも上手になります。残り30種類はパズルや冒険など、記憶や思考を必要とするものです。

いまの技術が出てきた2013年時点でこれをロボットに適用すると、ロボットがだんだん上達するということがほぼ自明だったのですが、実際にやったのがUC Berkeleyで、昨年5月です。このロボットが、オモチャの飛行機の部品を本体に組み付けています。強化学習でやっているのが最初は下手ですが、だんだん上達してきます。

上からのカメラ画像入力なので、ブロック崩しと原理的には全く同じで、上手にできたら報酬が貰え、この報酬を最大化するという仕組みで上達していくわけです。見てのとおり子供っぽいやり方で、適当に当てて入りそうなら押し込んでいます。レゴブロックをくっつけることもできるようになっています。これも、位置と角度の両方が合わないといけないのでかなり難しい問題です。

報酬を与え状況を変えると、いろいろな動作を学習させることができます。

いまと全く同じディープラーニングと強化学習を使って運転を学習させたのが、日本のプリファードネットワークス社とトヨタが共同でやって、今年のCESという展示会で出したものです。白い車が、強化学習で運転が上手になった車で、信号がない交差点を上手に通ります。赤い車は人間が運転していて、めちゃくちゃな運転をしますが、白い車は上手によけます。なぜかという、白い車は下手な状態から練習して上達しているのです。どうやったらぶつかるか、よけられるかがわかっているのです。人間の運転手と同じように、一旦上手になるといろいろな状況に上手に対応できます。車同士で通信していないのに、すごく上手に通り返けて行くことができます。

これも全く同じですが、ディープラーニングと強化学習をロボットアームに適用するというのを今年3月にグーグルが出しました。ロボットアームを14台並列に並べて、いろいろな物が入った箱の中から、目的のものを取り出す練習をさせています。経験を共有できるので14倍高速に学習できるということです。訓練により、いろいろな物を上手に掴めるようになります。

「これ、何かすごいのか？」と思うかもしれませんが、いままでの機械、ロボットは、こんな簡単なことすらできませんでした。まず目的のものを見つけるのは認識の問題なのでいままでは上手にできなかつた。しかも、物によって持つべき場所が違って、長いもの、丸い物、軟らかい物は、それぞれ持ち方が違います。

いままでのやり方は「こういう物は、ここを持って」と人間が定義していましたが、たかさんの物が混ざっていると持てませんでした。ところが、これは上手に持っている。なぜかという、何度も持っては落とし、練習しているからです。だから、一旦上達すると何

でも持てるようになる。人間の赤ちゃんもロボットも全く同じことです。

人工知能の世界では**モラベックのパラドックス**というのが 1980 年代から知られていました。コンピュータにとっては、子供のできることほど難しい。大人なら、あるいは専門家ならできるような、例えば医療の診断などは人工知能にとっては案外簡単で、人工知能研究初期の 1960～1970 年代からかなりできていました。ところが、3 歳児でもできる画像認識とか積み木積みが一向にできるようにならなかった。これは逆説的なので、パラドックスと言われていました。いまそれが覆りつつある。これが今回の第 3 次 AI ブームの本質だと思います。

福島邦彦先生が 1980 年ぐらいに提唱していたことと、いまのディープラーニングの仕組みはほぼ一緒ですが、いまは計算機のパワーがどんどん上がってきて、ようやく認識精度が人間を超えるところまで来たということです。

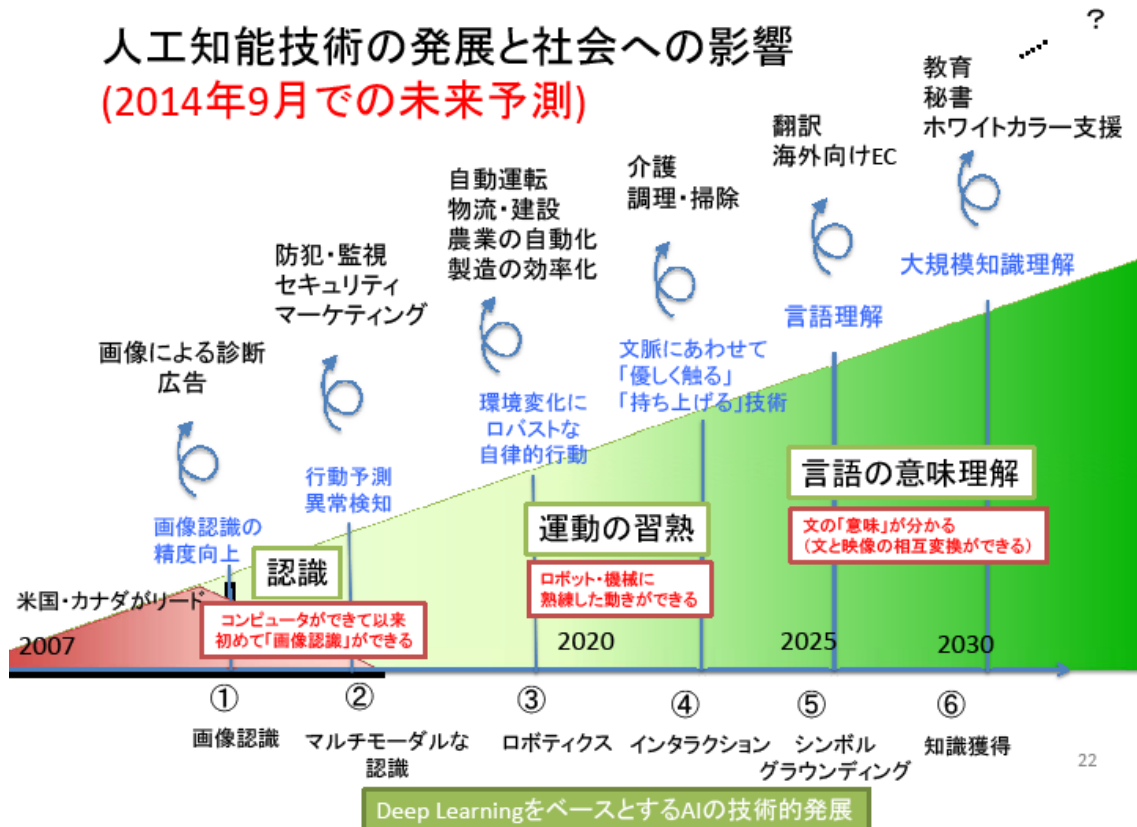
人工知能の研究は、「人間の知能はコンピュータで実現できる」という仮説からスタートしています。人間の脳は電気回路で、その働きは情報処理だとしたら、すべての情報処理はコンピュータで実現できるということをアラン・チューリングという人が万能チューリングマシンという概念で言いました。ところが、人間の脳がやっていることが 60 年経ってもできなかったのは、認識の計算量が非常に高かったからです。いまそれが乗り越えられつつあるので、元の仮説に戻ったほうがいいという時代に差し掛かっています。

ディープラーニングで認識、運動、言語が順番にできるようになります。これは赤ちゃんの発達過程と似ています。いまやコンピュータの認識は、静止画に関しては人間の精度を超えるところまで来ています。それから、運動も非常に簡単な動作であれば上手にできるようになってきています。言語とは言葉の意味理解をするということで、いまは自然言語処理によって言葉を扱うことができます。いまある技術は言葉の意味理解はしていません。英語をグーグル翻訳で日本語にするときには、ある英語の文がある日本語の文に置き換わる確率を統計的に計算して選んでいるだけです。

文の意味がわかるということは、昔から議論がありますが、単純化して言うと文から画像や映像を生成することで、つまり、思い浮かべることです。逆に、画像や映像から文を生成すること、つまり、考えたこと、思い浮かべたことを言葉にする。こういう文と画像・映像の相互変換ができることが、言葉の意味がわかるということだろうと考えますが、それが技術的に可能になると思います。

いまお話したようなことが 2030 年ぐらいまでに起こると思って描いたのが、この図です (図 1)。

# 人工知能技術の発展と社会への影響 (2014年9月での未来予測)



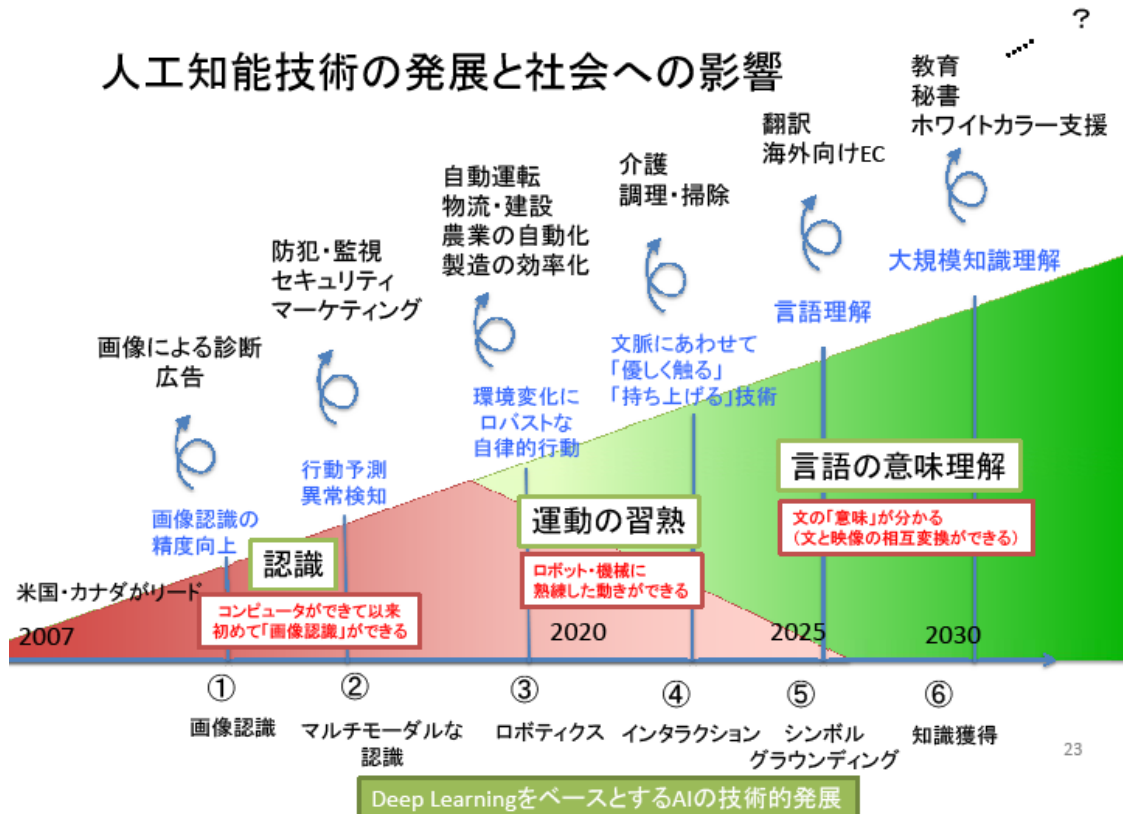
(図 1)

技術の進展とともに産業や社会に大きな影響があるだろう。例えば、認識ができるようになると、医者が X 線写真を見て影を見つけることは認識の問題なので、自動でできるはずです。

それから、防犯・監視。例えば警備員による見張りや、防犯カメラのモニター監視は認識の問題なので、全部自動化できるはず。そして、運動の習熟まで行くと、自動運転だけではなく物流、建設、農業の自動化。介護、調理、掃除なども自動化できるはず。更に言語の意味理解までいくと、翻訳、ホワイトカラー支援の全般ができるはずだと思って描いたのが、この図です (図 2)。



# 人工知能技術の発展と社会への影響



(図 2)

これは 2 年前に、当時、論文が出ていた分野を赤の三角形で書いています。2 年経つと三角形はこの辺まで来ています。画像認識という一番の峠を超えるとあとはすごい速さで技術が進んでいます。

言葉の意味理解というところにやや差し掛かりつつあります。これはどういうことかという、1 つは、Automated Image Captioning、自動画像キャプションづけという技術が出てきていて、画像を入れると文が出てきます。男性がギターを弾いている写真を入れると「man in black shirt is playing guitar.」という英語の文が出てきます。

更に、昨年末からは逆が可能になってきて、文を入れると絵が出てくる。例えば、「A very large commercial plane flying in blue skies.」と入れると、飛行機が青空を飛んでいるような絵が出てきます。「blue」を「rainy」に変えると、雨っぽい絵に変わります。これは、画像検索ではなくて描いているのです。ですから、例えば、「A stop sign flying in blue skies.」、「停止標識が青空を飛んでいる」という文を入れると、停止標識が青空を飛んでいるような絵を描きます。これは我々が情景を思い浮かべることとすごく近いことができるようになってきているということです。

ここまでできると、この絵から先ほどの自動画像キャプションづけで日本語の文に直すことが、技術的にはもう可能なわけです。結局、英語から日本語への翻訳になっているわけですが、一旦画像を介しているのの意味がわかって訳しているわけです。

もちろん、静止画を動画にしないといけないし、あるいは抽象的な概念をどうやって表すのかとかいろいろ問題はありますが、こういう方式によって、10~15年ぐらいで本当の意味での自動翻訳が実現すると思います。「ドラえもん」に登場する「翻訳コンニャク」という道具は、日本語でしゃべると英語で出て来る。それと同じようなものが出来る。英語で聞いても日本語で聞こえるようなものが出てくるはずです。

そうすると、言語の障壁が非常に低くなるので、日本の社会、経済は大きく変化すると思います。海外に販路が拡大すると思います。逆に、日本語の障壁が一気になくなると、優秀な高校生は海外に行ってしまう、大学は相当ダメージを受けるはずで、金融やメディアも含め、いろいろな産業が大きく変化すると思います。

結局、人間も動物なので現実世界を観測し、パターン処理をして行動しています。人間がすごいのは、その上に記号のシステム、言語のシステムを乗せていることです。例えば、「風船が飛んでいる」という言葉を聞いて、その状況を思い浮かべることができるし、「風船を割ったらどうなるだろう」と仮定の思考をすることもできる。つまり、記号の空間とパターン空間を行ったり来たりしながら思考しているところが、他の動物と圧倒的に違うところです。更に、その上に、この言語を他者とのコミュニケーションに使ったり、知識の蓄積に使ったりして知能のレベルを上げているわけです。

いままでの人工知能は、パターン空間での処理というのは非常に弱かったので認識すらできなかった。ところが、いまはディープラーニングでパターン空間の処理、つまり認識したり、予想したりすることができるようになってきました。そうすると、ようやく知能の研究の中で一通りの道具が揃います。

僕は、ディープラーニングが万能だと言うつもりはありませんが、例えば、ディープラーニングと現実世界でのインタラクションを上手に組み合わせると、現実的な環境の中で上手に動けるような機械やロボットができます。あるいは、パターン空間での処理と、記号空間での探索を組み合わせると、アルファ碁のような囲碁でトッププロに勝つことができるということで、このディープラーニングとさまざまな技術が組み合わさって進展していくのだろうと思っています。

#### 「大人の人工知能」「子供の人工知能」

いま、「人工知能」というキーワードが氾濫しているので、僕は「大人の人工知能」と「子供の人工知能」と、2つに分けています。子供の人工知能というのは、子供でもできるような認識、運動の習熟、言語の意味理解ができるようになってきた一連の技術を指します。

一方で、IoTとかビッグデータ全般で、データが取りにくかった領域でデータが取れるようになってきました。この領域に対して昔からある人工知能の技術を上手に使うと面白いことができるというのが大人の人工知能です。これは、裏で人間がモデル化をすごく頑張っているんで、例外に弱いとか、変化に弱いとか、いろいろ問題はありますが、一見するとすごく難しいことをやっているように思うわけです。

もともと、グーグルなどの検索、e コマース、SNS 等、インターネットの世界と非常に

相性が良かったのですが、今後は医療、金融、教育などの分野では、大人の人工知能がどんどん発展すると思います。

他方で、**子供の人工知能**が発展していく分野は、認識するとか、運動の習熟するようになるので、現場、実世界に近いところです。具体的には、僕がいつも 3 大分野に挙げている**農業、建設、食品加工**です。これらの産業は自然物を扱っているので、自動化が困難でした。例えば、トマトは市場規模も大きいし、収穫に工数もかかりますが、未だに収穫ロボットがありません。理由は単純で、いままではトマトが見えなかったし、上手にもぎ取ることができなかったのです。ところが、子供の人工知能によって上手に認識できて、運動の習熟ができると技術的にはトマト収穫ロボットができるはずですが、農業というのは人手がすごくかかりますが、結局、人が目の代わりをして、認識の仕事をしているわけです。

間引きや、花とか実を落とす摘果・摘蕾の非常に工数がかかる作業も、同様です。そう考えると、農業の世界ではほぼすべての作業が今後は自動化できるはずですが、ほとんど人手でやっている今の状態から、ほぼ全自動になる状態までの変化は非常に大きいと思います。

建設も同じで、建設現場の人たちは、認識しながら鉄筋を組んだり、コンクリートを固めたりしています。機械が認識できるようになると、建設現場も将来的にはほぼ全自動化できるはずですが、これも、いまの状態から全自動になる変化はすごく大きい。

それから食品加工、調理です。人が調理するのを当たり前だと思っていますが、それしかソリューションがないからです。調理というのは、典型的に認識や運動の習熟を必要とする仕事なので、機械にやらせられないのです。ところが、ディープラーニングによって、調理は自動化できるようになります。そうすると、いずれはほぼすべての飲食店の裏側は機械化するわけです。恐らく、集合住宅の中にもそういうのが入るでしょう。

更に、日本の食のレベルは非常に高いですが、海外に店を出すと、だんだん味が落ちてくる。これが海外進出の 1 つの壁になっていますが、バックヤードが機械化すると、日本の食の品質のまま世界中の飲食店のバックヤードを取れるので、産業としては相当大きいはずですが、日本人の食というのは非常にレベルが高くて、中華料理もフレンチも日本に来ると全部美味しくなる。それを機械に乗せて全世界に展開できると、自動車産業以上になるのではないかと考えています。

変化の本質は非常に単純で、画像認識ができないから人間がやっているたくさんの仕事が自動化されるということです。コストは 1/100 ぐらいになる。それから、運動の習熟と言うと機械も習熟するし、ロボットも上達するようになります。自然物を相手にしている農業、建設、食品加工でも自動化できるようになります。

更に、日常生活の中でもいろいろできるようになる。例えば、**片付け**です。片付けも、認識が必要なので、人がやるしかありません。ホコリはルンバを走らせておけばそれなりにきれいになりますが、片付けは認識がないと絶対できません。毎回散らかり方は違うのですが、今後は認識がディープラーニングでできるようになるので機械化できるわけです。

家でも、オフィスでも、商業施設でも片付けをします。全人類が一日で片付けにかけて

いる総時間は、恐らく運転している時間の比ではないぐらい長い。片付けロボットができると、ホテル暮らしと同じぐらい生活感が向上するはずです。家事労働が軽減されて女性活躍の土台になると考えると、片付けの市場規模は非常に大きいと思います。日本の家電メーカーも全部片付けマシンだけやれば良いと思うくらい、非常に大きい産業が今後どんどん立ち上がってくると思います。

日常生活や生産の中で人工知能、ロボットがどんどん使われる時代が早晚来ると思っており、僕は、これを「決勝リーグ」と言っています。そこに至るまでに予選リーグ A、B がある。予選リーグ A は**情報路線**で、人を助けてくれる人工知能で、例えば、メール管理やスケジュール管理をしてくれる秘書のような役割。この路線はグーグル、フェイスブック、マイクロソフト、アップル、アマゾンが強くて、日本企業が割り込む隙がない。もともと、こういう情報路線は英語圏でやったほうが絶対強いので日本企業が勝てるわけがない。

予選リーグ B は**運動路線**で、物を動かしたり、加工したり、操作したり、調理したり、掃除をしたりする、メイドのような役割。この路線で決勝進出を決めている企業は、世界的に見てもまだないと思います。有力なのは自動車メーカー、産業用ロボット、建設用機械、農業用機械、家電メーカー、医療機器等だと思います。グローバルなシェアを持っている大きな日本企業がこういう認識や運動の習熟という技術を使ってシェアを伸ばしていくことはあり得ると思っています。ですから、予選リーグ B を日本企業が勝ち上がることは不可能ではない。そのうえで、決勝リーグでグーグルやフェイスブックと戦えばいい。そのほうが、よっぽど勝つ可能性があると思っています。

いまお話したことは、**眼の誕生**ということ。今から 5 億 4,200 万年前～5 億 3,000 万年前という非常に短いカンブリア期に、現存するすべての生物種が出揃いました。なぜかは、諸説ありますが、10 年ぐらい前にアンドリュー・パーカーが、眼の誕生が原因だとする「光スイッチ説」を唱えました。それまでの生物には高度な眼がなかったので、モソツと動くしかなかった。ところが、眼ができると捕食の確率が非常に上がり、逃げるほうも、早く逃げよう、隠れよう、いろいろな戦略が生まれるので、眼の誕生によって生物の多様性が一気に花開いたという説です。

ディープラーニングで認識ができるようになったということは、眼ができたということなので、いままでの眼がなかった機械が今後カンブリア爆発を迎えて一気に多様化すると思います。昔からカメラはありましたが、イメージセンサーは網膜で、ディープラーニングは**視覚野**です。いままでは視覚野がなかったので、防犯カメラで撮っても結局人間が見ないといけなかった。ところが、いまディープラーニングで視覚野もできたので、ようやく見えるようになってきたのです。このカンブリア爆発を迎える、眼を持った機械を日本企業が取れるのかどうか最大の焦点で、今後数年の日本の未来を占う時期だと思います。

眼を持った機械ができることは、たくさんあります。例えば、警備、防犯。介護施設や病院で、何か異状がないか見守る。交通違反の監視などの社会インフラ。あるいは、顔認証。お得意様には割り引くとか、社員は通す。それから、意匠の類似を判定するとか、表

情を読み取るとか、国家の安全保障、入国管理、警察業務。あるいは、実世界の最適化。店舗内のお客さんの行動を見て、店舗を最適化する。あるいは、防災。河川・火山をずっと見張る。あるいは、医療画像の X 線、心電図などを見る。こういうことが、全部認識だけでできるようになるはずで。

更に、運動系の機械、ロボットと組み合わせると建設現場での重機、掘削や揚重を上手にやる。あるいは、セメント固め、溶接、運搬、取付け等を自動でやる。あるいは、農業では収穫、選果、防除、摘花・摘果。あるいは、自動操縦、自動運転。産業用ロボット、特に組立加工。調理。ペットロボット。医療、介護、バイオ。廃炉。この辺もどんどん自動化できるはずで、どの 1 つを取っても相当市場規模が大きいと思います。

ただ、海外企業は非常に動きがいいです。英国の**ディープマインド社**を筆頭に、エンリティック社は医療画像のディープラーニングを使って診断していて、既に人間の医者より 1.5 倍の精度が高くなっています。アフェクティブ社は、映画やテレビ番組のどこで視聴者の表情が変わったかを読み取って広告を配置したり、コンテンツを良くしたりする。あるいは、衛星画像の分析、ホームセキュリティ等、どれも、数十億円規模で資金調達をしています。

更に海外の大手企業、グーグル、フェイスブック、マイクロソフト、アップル、アマゾンには当然いろいろやっています。製造業でも、GE がディープラーニングによる医療画像の診断をやり始めています。ダイソンが掃除機に眼をつけようとしている。ドイツのクーカ、オランダのモービルアイは車用の画像認識。韓国の LG、サムソンもディープラーニングを活用しようとしています。日本も急がないといけない。

始めること自体はそんなに難しくなくて、ディープラーニングの技術が注目され始めたのが 2012 年頃なので、ディープラーニングで世界トップの人ですらキャリアでは 3~4 年ぐらい。いま多分、世界中で一番強いのがディープラーニングをテーマにして博士号、Ph.D. を取った 28 歳ぐらいの人です。それより年上になるとディープラーニングよりも前の技術を勉強しているので、あまり関係ないわけです。

どんどん環境が整備されていて、ライブラリも揃っているし、教科書も揃ってきている。国際会議の新しい論文も、グーグル等がすごいお金で雇ってきた研究者がどんどん公開しています。ソースコードもどんどん公開されていて、しかも商用利用可能な状態で公開されているので非常に入りやすい。数学とプログラミングが必要なので、理系の真面目な人向きです。そういう人が 3 ヶ月から半年もやれば結構できるようになる領域です。

一口にディープラーニングと言っても、3 つぐらい系統があります。1 つは **CNN (Convolutional Neural Network)** で、これは画像認識でいまデファクトになっています、視覚野の部分に相当する技術です。

2 つ目は **RNN (Recurrent Neural Network)** で、時系列の処理が得意なタイプです。

3 つ目は**オートエンコーダ**で、生成ができる。絵を描いたり、次のシーンを予測したりするもの。この 3 つがあって、目的に応じて使い分けるといことです。

人材の獲得については、国内で技術力があるベンチャーは限られています。

グローバルに見ても人材が足りないのが、この技術が持つ潜在的なインパクトを考え合わせると、社内育成したほうが良いと思います。一番若い理系の人間にディープラーニングを勉強させて何ヵ月かするとちょっとできるようになり、そういう仲間が2人、3人と増えてくるとだんだん強いチームになってくるという感じだと思います。結構真面目な技術なので、ある程度理解した後は地道にチューニングや工夫をしていくという日本人向きなところもあります。興味を持っている若者が、社内でこっそりやっていたりするので、それを励ましてあげて「もっと大っぴらにやっていたいよ」という感じかなと思います。

製品を設計するのではなく学習をさせる。この考え方の違いはかなり大きくて、いろいろな企業でも、どうしても動作を自分の頭で理解したい方が多い。ディープラーニングは、学習なのでたくさんデータを与えると学習してしまうわけです。例えば、人間のトラック運転手がどうやって運転しているのかを完全には理解できないのと同じで、そういうものだと思うしかないわけです。

どのようにデータを与えて、学習させていくかというのはノウハウのところ、例えば、トマト収穫ロボットには、いろいろなトマトを上手に見分けられるように、データを用意しないといけない。それぞれの領域でノウハウも違うしデータも違うので、その辺は、ノウハウを溜めていくしかありません。

経営的な側面から言うと、認識や運動の習熟で何ができるようになるか。日常生活や仕事の場面の中で、機械が認識できないからこういうやり方をしているというところがたくさんあり、それを見つけていかないと、どのように変わっていくかがうまく読めません。

機械が眼を持ち、認識や運動の習熟ができるようになるということは非常に大きな変化です。人工知能の話をする、「業務の効率化で経費を5%削減したい」というようなご相談も多いですが、そうではなくて、どんな会社でも売上が5倍、10倍になるかもしれないというくらいの大変化なので、一番大きな変化から真っ先に考えるべきだと思います。

例えば、眼を持ったこういう機械をつくるとこのマーケットが取れます。その次に、このマーケットが取れます。そういう戦略を考えて取りに行く。例えば、飲食の調理が全部機械化されるというときも牛丼屋から攻めるのか、あるいは家庭用の調理器具からなのか、いろいろな入り方があるはず。どこから攻めて、どういうマーケットを取っていくのかという事業戦略をしっかりと考えて投資することが重要だと思います。

最終的に、ディープラーニングは技術としてはコモディティ化するので、最終的な勝負はデータ量とハードウェアになります。

データについては完全に先行者有利で、トマトの収穫ロボットを一刻も早く投入したほうがデータは溜まるので、トマトの判定精度が良くなる。そうすると、製品の品質が良いからトマト収穫ロボットがより売れるわけです。そのように、ポジティブなフィードバックでどこも追従できなくなる。インターネットの世界で起こったことと同じことがいろいろな領域で起こっていく。

そして、再投資のループを作ったほうが勝つと思っています。ですから、トマト収穫ロボット、トマトの精度命みたいな会社は、トマトの収穫精度を上げることに一所懸命になるのでどこよりも精度が上がる。それで出た利益を再投資するから、他社はずついて行けなくなる。そういう再投資のサイクルを作れるかどうか重要です。

これは、社内文化との戦いという面も大きい。日本の大企業では、「うちで一番偉いのはエンジンをつくっている人だ」という場合が多い。情報システムは最下層。機械学習は情報システムの技術なので、この人たちがエンジンをつくっている人と同等以上に付加価値を生み出すのだという社内の合意を作らないといけない。恐らく、これはトップダウンでやるしかありませんが、かなり重要です。ですから、「ディープラーニングを使えばいいのに」と思う企業はたくさんありますが、社内事情で踏み出せない場合も多いです。

僕は1つのアイデアとして、視覚野の部分、頭脳の部分の最終製品、つまり「学習済みのモデル」をつくる**学習工場**をつくらばいいと思っています。頭脳の部分をつくら、それを最終製品にコピーして売る。その頭脳の部分をつくるには何が必要かという、**人とデータと計算機**の3つです。

あのトヨタですら、シリコンバレーに5年で1,000億円の研究所をつくらったというのが、多分、日本でも最大の投資です。しかし、研究所ではダメで、最終製品をつくらないといけない。それには当然、最終商品として何をつくるのかを明確にしないといけないし、事業計画を決めなければいけない。

日本の製造業は全然本気を出していません。本気を出したら、数十億円から数千億円という投資規模になるはず。そのぐらいでやっけて初めてシリコンバレー、グーグルやフェイスブックと勝負できると思っています。

そのためには本気で最終製品をつくらないといけないので、工場という言い方をしています。工場なので、10倍の生産をしようと思えば10倍の投資規模にならないといけない。

学習用の**データ**を作るにはどうしたらいいか。例えば、トマト収穫ロボットをつくるためにはトマトの画像をたくさん撮らないといけないし、採っていいトマト、よくないトマトというのをたくさんアノテーション、タグづけをしないといけない。こういうデータ作りは非常に大変で、人手もかかります。ここが大量の雇用も生むし、2次サプライヤー、3次サプライヤーもできるような世界ではないかと思っています。

次の**計算機**の部分は**エヌビディア社**が強いので、これは買えばいい。海外の研究機関も非常に強いので、これは地道に真似をすればいい。

人については高度なスキルを持った人材が圧倒的に足りないですが、社内でも育成したほうがいいと思うし、大学でも育成していくということだと思います。

「眼のある機械」が1ついいのは、「モノ売りからの脱却へ」と書いていますが、必ずデータが自社のサーバを通る形になります。例えば、トマト収穫ロボットで精度を上げたいと思えば、トマト収穫ロボットが認識した画像は必ず自社のサーバに上げるような構造になる。トマトを認識する技術はどんどん上がっていくわけです。トマトの収穫の他に、病

気のトマトを発見できるという付加価値がつくと、病気のトマト発見分に対してチャージできるわけです。ですから、サービスに対してチャージをするということが自然にできるようになる。しかも、認識できるもの、動作できるものが増えるほどチャージできる額が増えていくので、グーグルやアマゾンと同じようなサービス型のモデルが自然にできます。

最終的には、畑、出荷計画、全部含めてサービス化していけばいいので、日本が強い 1 つの製品を基点として、ネットワークに繋げてサービス化していくことが王道ではないかと思っています。

少し社会全般の話をしてしまうと、よく言われるのが、例えば、自動運転で A さんを轢くのか、B さんを轢くのかという**トロッコ問題**、**トロリー問題**が起こった場合にどうするのか、事故の責任は誰になるのかという議論があります。更に、人工知能が犯罪に使われたり、軍事に利用されたりする可能性がある。眼を持った機械なら人の顔を見つけたら撃つという兵器活用は非常にやりやすいので、国際的な議論が必要です。あるいは、人工知能が創作した絵や楽曲の知財はどうなるのか。あるいは、認識能力が上がると防犯・監視レベルが非常に上がるので治安は良くなりますが、「見られない権利」の議論も非常に重要です。

よく人工知能が人間を襲うのではないかと言われますが、そういうことはありません。なぜかというと、生命と知能というのは別物だからです。知能というのは、目的が与えられたときの問題解決の力のことです。生命というのは、生き残りたい、子孫を残したい、仲間を助けたいというような目的を持っている。人間の場合は、生命としての目的を、知能を使って上手に達成しているということです。人工知能というのは知能の技術なので、基本的には問題解決の技術です。それが、ある一定以上レベルが上がると急に生命的な欲求を持ち始めるということはありません。

ただ、目的の設定が非常に大事で、悪い人が人工知能を使うことは大いにあり得えます。社会の中で利害が対立するときに、どちらがどのぐらい譲歩すべきかは人間が決めないといけないことで、人工知能の技術が進展するほど、社会の中でどういう価値観を上位に置くのか、そこを、問い直さないといけません。

日本のいろいろな社会課題に対して、眼を持った機械、ディープラーニングと物づくりを掛け合わせたものを使っていくことはいい戦略だと思います。例えば、農業分野では、休耕地が耕せるし、農家の収量が増える。介護では介助も楽になるし、高齢者も自立した生活ができるようになる。廃炉の工期を短縮できる。河川や火山を見張って、防災ができる。しかも、こういう技術を伸ばして、新たな輸出産業にしていける可能性があります。

眼を持った機械は日本にとって大きなチャンスです。日本は物づくりが非常に強いし、それを支える文化的な背景や、いろいろな産業の広がりもあって、これは大きな強みです。ディープラーニング技術は物づくりと非常に相性が良く、眼と機械・ロボットの部分を掛け合わせることで相当いろいろなことができるわけです。

そういう意味では、いま日本は非常に大きなチャンスを迎えています。これは競争なので勝たないといけない。イノベーションの速度では日本は遅いので、僕はもう 2 年ぐら



い講演でこのことを言い続けています。何とか、動いてほしいです。やはり、正しく速く動いていくことが重要だと思いますが、1つには人材育成していく。社内でも育成していくということです。それから、経営的な側面からは事業や産業がどう変わるのかを見抜いていくということ。社会全体では、新しい未来像を描いていくことが重要だと思っています。ご清聴、ありがとうございました。

(平成 28 年 11 月 7 日 (月) KKR ホテル東京にて)



RESONA

## **リそな中小企業振興財団**

**The Resona Foundation  
For Small And Medium Enterprise Promotion**

〒141-0021

東京都品川区上大崎三丁目 2 番 1 号

Tel. 03-3444-9541 Fax. 03-3444-9546

URL: <http://www.resona-fdn.or.jp>

E-mail: [staff@resona-fdn.or.jp](mailto:staff@resona-fdn.or.jp)