高知女子大学看護学会誌　VOL. 47, NO. 1, pp103～122　2021

第47回 高知女子大学看護学会 講演会

人の心とAI

京都大学大学院情報学研究科知能情報学専攻脳認知科学講座　教授

熊田　孝恒

私は、元々は大学院では心理学の勉強しておりました。その後は研究所等で、いろいろな企業の方と共同研究をしてきました。しばらく前から自動車メーカーの方々と共同で自動運転の研究に関わったときに、自動運転そのものではなく、自動運転の人工知能（AI）を搭載した車と人がどう関わり合っていくのかという研究を始めました。その頃からAIと人間の関係を調べる研究をいろいろ行ってきています。現在は、ロボットを実際に作っている人たちと共同で、人を助けるようなロボット、特に最終的には鉄腕アトムのような人と協調しながらいろいろな問題を解決していくようなロボットを作りたいと思っていますが、ロボットそのものを作る技術というよりは、そういうものを作るために人のどういう側面を我々は知っておかないといけないのか、そういうロボットと人との関係の中でロボットはどのようなことを理解して、どのような振る舞いをしなければいけないのか、というようなことを調べたりしています。今日の話もテクノロジーの話もさることながら、ロボットではなかなか実現できないロボットの側面、特に皆さんが日頃関わっておられる看護のような世界にある、ロボットではなかなか実現できないことが、人間のどんな能力と関係しているのかという話を中心にしていきたいと思っています。

■背景と本日のあらすじ

まず、今日の話ですが、最初にAIとはどのようなものかという話を簡単にさせていただきます。いろいろな定義がありますが、広く言うとAIというのは人間の行う知的な活動の一部と同じようなことを行うコンピュータ上のプログラムのことです。知的な活動と言っても幅広いのですが、最近は様々な応用が期待されており、また、かなり実用的なものも増えてきています。特に、AIの技術の進展というのは、ここ数年で著しくなっています。背景としてはコンピュータそのものの処理速度、あるいは処理能力が上がってきたということがあり、それと相まってAIの技術そのものも進歩しています。AIは、医療を含めていろいろな場面で、今後、我々の生活を大きく変えていく可能性があります。その辺の話をしながら、ただその延長線上に、最終的には人の心に寄り添うとか、人のように振る舞うとか、あるいは人を理解するといったことがAIに期待されているのですが、そんなことが本当に可能なのかということを、人間の心ってそもそもどんなものかという観点から簡単に皆さんと考え、可能性みたいなものを少し検討したいと思っています。最後に、今後、我々とAIの付き合いはどうなっていくのかということをまとめてお話を終わりにしようと思っています。

■弱いAIと強いAI

AIは人工知能artificial intelligenceの略で、日常用語でも使われています。AIというものを考えるときに、1980年頃が第二次AIブームと言われていて、そのころも万能なAIができるんじゃないかと言われて、研究が進んだ時期もありました。その後、やっぱりそんなの無理じゃないかという時期があって、一旦下火になり、今が、第三次のAIブームだと言われています。

第二次のブームの頃に、弱いAIと強いAIというのを分けて考えようと、哲学者のサールという人が言い始めました。弱いAIがどういうものかというと、限られた範囲の知的な作業を行う

プログラムです。決まった限られたことにはすごく力を発揮するけど、それ以外だと全くできないというようなプログラムのことを弱いAIと呼びます。一方、強いAIというのは広範な知識を持っていて、人間と同様に、あるいはそれ以上に様々な問題を解決できるものです。鉄腕アトムのようなものは、我々が子供の頃にイメージしたAIを搭載したロボットですが、いろいろなことに対応できて、未知のことでも解決してくれる。おそらく意思だとか計画性だとか、今人間にしかないであろう幅広い固有の機能をAIロボットが持つようにならないと、こういうことは実現できないのですが、我々がイメージするような将来的な強いAIというのは、こういうAIのことです。

最近、汎用人工知能artificial general intelligenceということが言われていて、何にでも役立つ人工知能を作っていこうということが考えられています。今の弱いAI、例えば自動運転を実現するようなプログラムは、自動運転には使えるけども、そのプログラムを全く別のことに使おうとすれば、また一から作り直さないければなりません。弱いAIについては、ものすごく研究開発されているのですが、一方で、強いAI、あるいは汎用AIをどう作るかというのが今後の大きなテーマになってきています。

最近、機械学習とかマシンラーニングという言葉をお聞きになった方がいるかもしれません。これが今のAI技術の中心的なものです。機械がいろいろなことを学習することによって、専門的な知識を利用して、ある能力を発揮できるというものです。この技術が進んだから今のAIがあるということになります。機械学習は、ビッグデータ、つまり膨大なデータをもとに、コンピュータがある特定のことをひたすら学習するというものです。その学習方法には、大きく分けると、教師あり学習と教師なし学習と言って、正解を教えながら学習させるのと、コンピュータそのものに正解を見つけさせるという方法の２種類あります。例えば、いろいろなところで画像診断、特に医療的な画像診断がAIで行われていて、人間のエキスパートと同じ程度の能力か、それ以上の能力を発揮できるということが言われています。その背後にあるのは、深層学習deep learningと呼ばれるような機械学習の手法です。これについても簡単にお話ししたいと思います。さらに、最近、敵対的生成ネットワークという新しい人工知能の考え方が出てきて、これがまた世の中を大きく変えうる可能性があるので、これも極々簡単に説明したいと思います。

■教師ありの機械学習

教師ありの学習がどのようなものかというと、正解がわかっていることについて、その正解を導くようにコンピュータが学習をするというものです。例えば、代表的な応用例としては医療画像があります。画像診断とは、皆さんご存知のとおり、画像の中に病変が含まれているかどうか、病変がどのような種類のものなのか、例えば、どれ位のグレードの腫瘍なのか、またどういった病名がつきうる状態なのか、などを診断することです。これは熟練した医師がこれまでして来た仕事です。医師は、これまでの経験の中から画像の様々な特徴や、その他の様々な情報を組み合わせて診断を下していたのだと思います。それを機械にやらせようという話です。病変を含む臓器の個々の画像に、熟練医師の見解や病理的な検査などをもとに、最終的な診断結果、つまり正解のラベルをつけていきます。診断がついている膨大な画像をもとに陰性か陽性かということを機械に学習させます。エキスパートの医師の脳の中には、例えば腫瘍の大きさと形などに関する知識や経験に基づくさまざまな情報が詰まっていて、それが多次元の空間に情報として表現されていると考えます。その中で、ある条件の組み合わせを満たしているものについては悪性度が高く、そうでないものは悪性度が低いと判断されます。つまり、エキスパートの頭の中の知識の多次元空間の中には、病気かどうかの境界線があると考え、その境界線を機械に学習させようというのが、教師ありの学習と呼ばれるものです。これは、名医の方がされているようなことを機械に覚えさせるということなのですが、現在では、十分なデータ、つまり十分な量の病変を含んだ画像のデータと診断結果があれば、いろいろな疾病に対して応用が可能になりつつあります。医療場面に限ら

ず、自動運転のような場面でも、何か異常を検知するとか、物を見てそれが歩行者かなどを判断したりといった、さまざまなものに利用されるようになってきています。ただ問題は、人間が判断しているのと同じ特徴を使ってAIが判断しているとは限らないという点にあり、今後、このことが問題になるかもしれません。

■深層学習deep learning

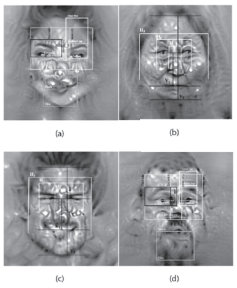
近年、deep learningという技術が開発され、複雑な機械学習が可能になってきています。これは誰かが手書きした０ ～ ９までの数字を判別させるような人工知能のプログラムを考えるときに、ニューラルネットワークの技術は有効です。入力の手書き文字の画像が与えられて、それが５段階位の処理を経て、最終的に０ ～ ９のどの数字にもっともらしいかという確率の計算をニューラルネットワークで実現できます。

deep learningのdeepは、何がdeepかというと、先ほどの５段階位のプロセスを経て、最終的に結果を導けるようなプログラムの５段階のところを、例えば100段階とか150段階というレベルで非常に深い、非常に複雑なレベルの情報を解析するような段階を設けられていることに由来します。その結果、より複雑な学習が可能となります。例えば囲碁の世界チャンピオンに勝つようなプログラムがdeep learningを使って実現できたわけです。

■説明可能（explainable）なAI

どういう画像が入ってきた時に、どのような出力を高めるのかといったパラメータの重みづけを変えていくのが機械学習と呼ばれるものです。しかし、最終的にどうしてこういう結果が導かれたかということを知るのはなかなか難しいのです。というのは、ある画像を入れたら「２」という数字が一番大きくなるような値が出てくるのですが、中のどんな関係によって、そのような結果が出てくるかを我々が知ることはなかなか難しい。100階層ぐらいあると、どこでどういう結果がどういうふうに働いて、こういう出力が出ているのかがよくわからないのです。ということで、このdeep learningに関しては、この答えを導く過程がブラックボックスだと言われています。つまり、ある医療画像を入力した時に、これは腫瘍のこういうものですということが、結果としてほぼ正しく出てくるとしても、なぜこの画像だとそういう診断になったのかということを、人間が知ることはなかなか難しいというわけです。AIそのものもいろいろな症例をひたすら学習していって、いろいろな組み合わせが起きたときには、それはこういう症例だと結論を導くように学習がなされます。結果は確かに正しいのですが、じゃあどういう組み合わせだからそういう症状に診断が下ったのかということを知るのは非常に難しいというのが、この深層学習の一つの弱点です。最近は、explainable（説明可能）なAIといって、処理の過程を説明できるように、可視化しながら学習させる方法が提案されています。後の方で説明しますが、最終的にAIが言うことを我々が信じることができるのかということに関しては、やっぱりどうしてそういう結論が導かれたのかということをちゃんと説明してくれないと、我々は納得できないのだと思います。だから、この中身を透明にして、どういうプロセスでこういう結論が導かれたのかを知ろうという取り組みが行われています。

Tong, Liang, Kumada & Iwaki (2020)の研究



- 画像とカテゴリ（男女、魅力的か否か）のラベルを与えて学習しただけで、中間層のニューロンが黄金比を表現するようになる。
- 顔のパーツの黄金比が魅力度と関係があることは、ダ・ビンチの頃から知られている。
- AIが、人間と似たような判断基準を自動的に獲得できることを示している。

我々の研究室でも、ニューラルネットワークが、結果を導く途中でどんなことが起きているのか知ろうという研究をしています。世の中に一般に公開されている魅力的な男性と女性の顔写真と、あんまり魅力的じゃない男性と女性の顔写真を使って、先ほどのdeep learningのプログラムを使って、それぞれの画像が男性であるか女性であるか、その人はどれぐらい魅力的かということを学習させます。そうすると、学習さ

せてない顔を入れたときに出てくる値と、元々その顔について人がどれ位魅力的かを評価した値を比べると、比較的相関が高くなるということが分かりました。つまり、こういうことをさせると、人が評価をするのと同じように、AIのプログラムも顔の評価というのを学習できることが分かります。このこと自体は、そういう学習をさせれば今のAIのプログラムであればそれぐらいのことができるということが分かっていたのでそんなに驚きではありません。問題は、このプログラムがどういう条件を使って、魅力的かそうでないかを評価しているかということです。これを先程のexplainableな方法、つまり、プログラムの中でどんなことが起こっているかというのを可視化する方法があります。それを使ってみると、左の上の図は魅力的な女性の顔をAIが判断する時の情報の種類を集約したものです。このような顔に近い顔をAIのプログラムは魅力的だと判断します。その下は魅力的な男性の顔のテンプレートです。こういう顔に近いと魅力的だとこのAIのプログラムは判断します。この右上は魅力的でない女性の顔のテンプレート、その下は魅力的でない男性の顔のテンプレートで、このAIのプログラムはこのような特徴を学習しているということです。

面白いのは、人間の顔の魅力度というのは、目と鼻の間の距離と、目の間隔のようなものが黄金比になっていると魅力的に感じるということがレオナルド・ダ・ヴィンチの時代から言われ、研究されてきていますが、まさにそういう黄金比みたいなものが、この魅力的な顔のテンプレートの上では表現されていて、魅力的ではない顔のテンプレートでは表現されていないことがわかります。実際には、魅力的な顔とそうでない顔を見せて、人はこういう顔の時には魅力的に感じ、こういう時には魅力的に感じないということをAIのプログラムに教えているだけなのですが、AIのプログラムは自動的にその顔の中から黄金比みたいなものを抽出して、そういうものを判断の基準として利用していることがこういう研究から分かってきたというわけです。AIのプログラムは、人がやっているのと同じように結果を導いているのかは分からないと言いましたが、中身を可視化していくと、人がやっていることと同じような方法で判断しているのか、あるいは全然違う方法で判断しているのかということが分かってきます。

■自動認識技術による誤認識

自動認識技術による誤認識

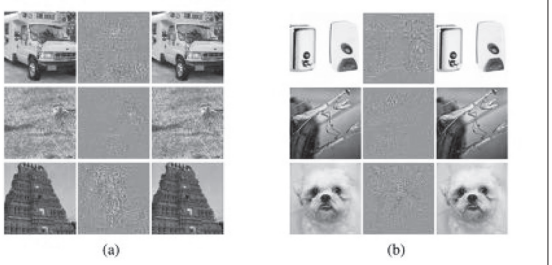
- 少し前から、SNSなどで画像認識機能を備えた某社の自動車で、ラーメンチェーン店の看板を侵入禁止の標識を誤認識するということが、話題になっている。
- 恐らく、学習の段階で、このような事例が入力されてないため。
- 未学習のものについては、正しく判定できるとは限らない。

侵入禁止の標識

ラーメンチェーンの看板
https://www.tenkaippin.co.jp

先ほどの教師ありの学習というのは、教えられたことに関しては、人間以上に正確な判断ができますが、教えていないことは全くわからないということが起きたり、間違えたりということが起きます。画像認識機能を備えた某社の自動車で、侵入禁止の標識を検知したら、「今侵入禁止の道に入りましたよ」とドライバーに教えるようなシステムがついています。そういうものが、似たような形をしているあるラーメンチェーンの看板にも反応して、「侵入禁止の道ですよ」と教えるということが、少し前からSNSとかで言われています。学習させる時、どういう形のものが侵入禁止のサインなのかということを機械学習で教えるのですが、その段階でこういう紛らわしいものはこういう場所にあって、例えば横に文字が同時についているのは進入禁止のマークではないよということも、偽陽性の例として教えておかないといけなかったのです。しかし、それがなされていないので、侵入禁止のマークだというふうに判断してしまうということが起きます。ですから、機械学習では、まずもって学習していないものに関しては、正しく判断できるとは限らないということです。医療画像の診断などに関しても、全く見たことがないような画像、あるいはただ単にノイズか何かで画面上にありえないものが写っているとかそのようなものがあったとする

と、それを正しく判定はできない。医師ならば、これはたまたまこういうものが写り込んだに違いないとか判断できるかもしれないけど、AIのプログラムではそういうことができません。教わってないことは全く何もできないというのがAIの一つの大きな特徴です。



人間とは異なる特徴を使って判断している。 Szegedy et al. (2013)

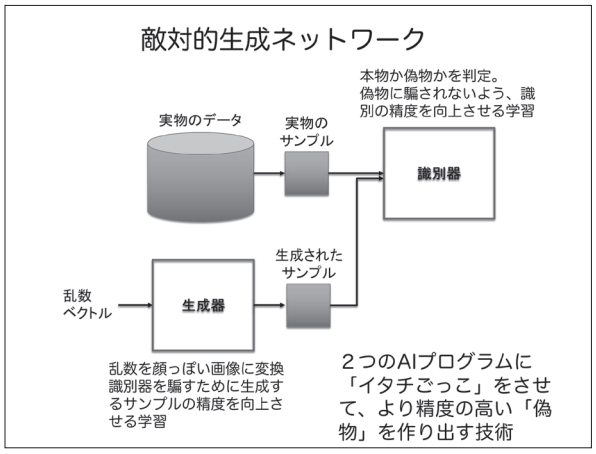
もう一つ、最近、AIの教師ありの学習というのは、我々と全く違う画像処理をしているかもしれないと言われています。元からそうかもと思われていたのですが、ある研究者たちがとってもドラスティックな、ものすごくそのことを感じさせるような報告をしました。ここにあるこの画像の左側の列、例えば左上のこれはトラックですが、このプログラムはこういう画像を見せたときにそれが何であるかを判定するということを学習しています。例えばこういうものを見たらトラックと答えるかもしれないし、ちょっと分かりにくいのですが、ここに七面鳥か何かが写っているとか。その下はビルか何かそういうふうに答えるようにできているようなプログラムです。ところが、このプログラムにちょっとだけノイズを加えます。左側の画像と右側の画像は、我々がみる限り全く同じもののようにしか見えないわけです。どう見ても同じトラックにしか我々には見えないのですが、実はこのプログラムは、右側の列はどれもダチョウと回答する。この真ん中の図は、この右の図と左の図で違いのあるところを色が違うように示しているのですが、ほぼ我々には感知されないような、真ん中にあるような多少のノイズを左側の画像に加えると、右側の画像はどれもAIのプログラムにはダチョウに見えるということになるそうです。我々にとってはそうは見えない。先程、教師あり学習では、境界線を学習するといいましたが、ダチョウとそうでないものの境界線をこのプログラムは学習しているはずですが、ちょっとしたノイズを加えることで、そのプログラムにとっては境界線を超えて、他の画像として認識されることを示しています。いずれにしても、我々には全く信じられないのですが、AIのプログラムというのは我々と全く違う情報を使って物事を判断しているらしいということが、これらの例からも分かってきます。

■教師なしの機械学習

機械学習の方法で学習したAIは、このように我々が予期しないような間違いを犯すこともあるようです。AIはとても優れているが、我々とは違う情報処理をしているが故に、我々とは違った処理特性を示す可能性があるというものです。一方、教師なしの学習というのもいろいろ盛んに行われています。これは、教師があるわけじゃないけど、ある評価の結果を最大化するように、AIが自動的に学習をしていくというものです。例えば、少し前に将棋のPONANZAというプログラムが、プロの棋士に勝ったという出来事がありました。その時に使われたプログラムは、ルールを知っているAIのプログラム同士を対戦させて、お互い相手に勝つように学習をしていきます。従来は、人間が指した手を正解として、それをひたすら学習して強いAIのプログラムが作られていたのですが、最近はプログラムのAI同士を戦わせて、どうやったら相手のプログラムに勝てるのか、ということを互いにひたすら学習していくことによって強くなっています。その結果、人間では太刀打ちできないレベルにまで達しているというわけです。「囲い」といって王様の守り方をAIのプログラムが発明したり、昔流行っていたけれど、勝ち目がないからと廃れた戦法が意外と実は強いらしいということがAIの発展によって分かってきたりしています。このように、人間だけではわからないような新しいことを、AIが発見していくということが起きてきています。

■敵対的生成ネットワーク

もう１つ最近のAIの革新的な技術の話をします。最近新聞にも載ったので、ご覧になった方もいるかとは思いますが、この左側にあるような顔で、「55歳男性安藤さん」とか、「49歳男性松本さん」という人がいるのですが、この人がいろいろな商品の感想を語っていたりという広告がwebサイトとかに載っています。実はこの安藤さんや松本さんは実在の人物ではありません。今、実在の人物ではないのだけれど、いかにもいそうな人の顔をAIが自動的にどんどん作ることができるというようになっています。



どうやって作っているのかというと、敵対的生成ネットワークといわれるような方法が用いられています。すごく簡単に説明した図がこれで、２つのAIのプログラムが存在します、１つは生成器というプログラムで、もう１つは識別器と呼ばれるプログラムです。生成器というAIは、乱数で作られた元々のすごくでたらめな画像からそれを変換して人っぽい顔を作り出します。最初は人らしいかどうか分からずに、ほぼランダムにどんどん作ってサンプルとして識別器に送るわけです。最初は本当にノイズと同じような全く顔の画像とも思えないようなものがやってきます。識別器は、送られてきたものが元々実物のデータの実物の顔なのかそうでないのかどうかを判定する機械です。識別器はいかに生成器からくる画像に騙されないようにするかといった識別する能力をどんどん学習して向上させていきます。それに対して、これは顔じゃないとダメ出しされた生成器は、今度こそということで、より識別器を騙せるように、その中のプログラムが学習していきます。その結果、生成器は識別器を騙せるように画像の精度を上げていくし、識別器は騙されないように識別の精度を上げていくということをします。つまり、この２つのAIがいたちごっこをすることによって、より精度の高い偽物を作り出すことができるということになります。先ほどの人の顔にそっくりな画像は、こういうプログラムを使っていかにも実在の人、つまり識別器に入っている膨大なサンプルと見分けがつかない情報を作り出すことによって作成されたものです。

このような技術を用いると、白黒写真に着色するとかも可能になります。これもAIは、最初はでたらめな着色をしますが、そのうちだんだん識別器の方が学習してくるので、でたらめな着色をしていたら、そんなのありえないとダメ出しをされ続けて、生成器はダメ出しをされないような着色をしていきます。あるいは線画に本物らしく色をつけるとか、普通の馬にシマウマっぽい模様をつけるとかというようなことができるようになってきます。つまり、AI同士がお互いに切磋琢磨というか、いたちごっこしながらどんどんお互いの精度を上げていって最終的に実際のものと区別がつかないものを作り出していくというようなことができるようになってきています。

■弱いAI

弱いAIの現状

- 正解がある問題に関して適切なデータが膨大にあれば、人間と同等、あるいは、それ以上の能力を発揮しうるだけの技術レベルに達している。
  - 質の良い、正解つきのビッグデータをどう集めるかは問題
- 学習していないことには対応できない。
- 人間とは異なる学習方法、判断基準を用いている可能性
- ある側面では、人間はかなわない。
  - 疲れない。
  - 忘れない。
  - うっかりミスをしない。

弱いAIと言われるような、あることに特化したAIというのは、正解がある問題に関しては適切で膨大なデータがあれば、人間と同等あるいはそれ以上の能力を発揮しうるだけの技術力

に達してきているということがいえます。問題は正解付きのビッグデータをどう集めるかっていうことですが、その問題さえ克服すれば、この先どんなものでもそこそこの精度で判定できるようになるのではないかと思われます。まさに医療画像の判断みたいなものは、すでに膨大なMRI画像だとか、レントゲンの画像だとかとエキスパートの先生方が診断された結果というデータが世の中に膨大にあるので、そういうものを学習させる事によって、本当に精度良く判断ができるようになってくることは確実だと思います。ただお話ししたように学習していないことに対応できないとか、人間とは異なる方法で学習しているので、人とは違う判断基準を用いている可能性があります。そういう問題点もあるものの、ある側面では人間は敵わなくなります。AIというのは疲れないので、24時間働き続けて学習し続けるし、レントゲン画像を見て、その中から病変を探すということを24時間だろうが何日間もひたすら眠らずに続けることができます。そして彼らは忘れないし、人がするようなうっかりミスもしないので、ある意味こういう点においては、人間は敵わなくなってきていることは確かです。人間とAIが間違う理由というのが異なっていて、AIは学習していないものは全く正しく判断できないし、ダチョウの例みたいに我々が予期しないようなノイズ、我々が感知できないようなノイズがあっても、間違えたりということが起こる可能性もあります。

一方で、人間はヒューマンエラーのような間違いを必ずするので、人間とAIを組み合わせることで精度が格段に上がるということが報告されています。例えば医療画像の診断なんかでは、AIの診断と人の診断を組み合わせることによって、お互いの弱点を補うことができて、精度が高いものになるということが、いろんなところで検証されています。

この先こういう弱いAI、ある事にだけをひたすら学習して、そこに特化したAIというのがどんどんできてくることになって、様々な応用が期待されています。一方、それによってなくなることが予想されている職業というのがいろいろ想定されています。例えば、会計士とか一般の事務とか秘書的な役割とかコールセンターとか受付とかなど、ある種入力のパターンが決まっていて、そこで起こりうる問題のパターンが決まっていて、それに対する対処の仕方も決まっているようなもので、ルール化していろんな事例をもとに学習が可能なような職業は、この先AIに置き換わっていくことが予想されています。看護だとか医療の中でも外科的な側面などは、最終的にAIではなかなか難しいのではないかと言われています。現在でも、外科手術もある部分はAIがサポートしたりということがありますが、AIが様々な症例に応じて様々な経験と技術を使って全面的に手術を行うというところまでには至っていません。看護の場面なども、パターンとして学習できないような様々な事例が起こりえるので、おそらくAIには置き換わらないだろうと言われています。看護は、最後にAIに置き換わるぐらいだと言われている職業なのですが、そういうものとAIで起き変わりそうと言われている仕事の違いをイメージしていただくと、この弱いAIがどこまで発展しそうかのイメージもできるのかもしれません。

■AIのひろがり

ここまでAIの技術の現状みたいなところをお話しさせていただきましたけど、ここからはこれが今後どうなるか、人間の生活に浸透していく上でどのようなことが問題になるのかということを、少し心理学の立場からお話ししたいと思います。

AIは、今どの辺まで広がってきているかというと、１つは弱いAIの世界では特定領域では人間をかなり支援できるようになってきていて、自動運転も然り、医療画像の診断というのもすごく進んできています。金融では、いろいろな株価などの動向などを判断して、この先どうなると予測をするようなことは、もうAIがほぼかなりの部分置き換わってやっています。世の中の金融為替市場というのも、かなりAIの力で動いているというのは確かです。また、最近、受付だとかそういう問い合わせ業務が、かなりの部分AIに置き換わりつつあります。自動翻訳の技術もかなり進んできていて、精度がとても良くなってきているのは確かです。特定の領域に関しては今後どんどんAIが進化していく、つま

り弱いAIが高度化していくことは確かです。

一方、強いAIの方、つまり何にでも役立つ汎用的な知能みたいなものの研究・開発はまだまだスタートしたばかりに近い。一つは、スマートスピーカーとかAIスピーカーと呼ばれる、スマートフォンについているSiriとかがそれに近いのですが、ちょっとした質問に答えてくれるとか、「注文して」といったらそれを注文してくれるとか、そのような感じのものはいろんなところに出てきています。また、ペットロボットみたいなもので、いろいろな人の好みだとか、いろいろなことを学習していって、その人と寄り添うというか、人の友達とまではいかないかもしれませんが、ちょっとしたパートナー的な役割をしてくれるものというのは、今後進展していくことが期待されています。

このように、強いAIに関する研究は、まだスタートしたばかりという感じです。おそらく我々が子供の頃に目にしたAIというのは、鉄腕アトムやドラえもんみたいに、人の気持ちを理解して、人間と同じように他の人を助けてくれたりするAIで、そういうものをイメージして夢を抱くのですが、そこについてはなかなかまだ実現は難しい。なぜ難しいかというと、そもそも人間がどのように世の中を理解しているのかとか、人の心をどう理解しているのかとかということがまだまだ十分に分かっていないからです。少なくともこれらがある程度解明できないと、強いAIというのは実現しないのではないかと私たちは思っています。一方では、そんなこと分からなくても作れるのだと言っている人もいますけれども、人間ってどうなっているのか、人間の気持ちに寄り添うには人間の気持ちがそもそもどんなものなのか、などが分かってないと作れないんじゃないかというのが我々の立場です。

■フレーム問題

人間が実世界で解いている問題

- 正解のある問題（＝弱いAIが得意な問題）ではなく、、

- 正解のない問題
- 正解が一意に決まらない問題
- 問題を解く前提が不明な問題
- 前提が多すぎる問題
  - 考慮すべき前提に関連する情報と、そうでない情報（ノイズ）が混在（＝フレーム問題）

人間を理解する上で、どういうところが難しいのかなということを考えていきたいと思います。皆さんが今日着ている服は、どのように選ばれたのでしょうか。あるいは最近外食もままならないですが、お昼ご飯をどこか食堂で食べたとかでもいいのですが、どうしてその食べ物を選んだのでしょうか。こういった我々が日常生活で解いている問題は、こういった類の問題です。つまり正解が必ずしもないですね。AI、特に弱いAIが得意なのは、正解がある問題を正しく解くということです。癌であるかそうでないかとか、そういうものを正しく判断するというのが彼らが得意な作業なんです。一方、我々人間が実世界で解いている問題、着る服を決めるのが大袈裟な問題じゃないかもしれないのだけど、ともかく正解がない問題というのを日頃から解いている。一意に決まらないとか、解く前提が不明な点が多すぎるとか、そういう問題を我々が解いてるんだということになります。朝着る服を決めるといっても、もちろん季節だとか、シチュエーションに合う服となると、そんなに候補が多くないかもしれないのだけど、皆さんのご家庭のタンスだとこの季節に着れる服でも10着ぐらいあるかもしれませんね。その中で、今日はこれこれこうだからと決めていって、最終的には着る服を決めてるのだとは思います。

そういう問題というのは、AIではなかなか解きにくい問題です。人もそういうことが途端にできなくなるという症例が報告されています。これは、前頭葉の両側の脳腫瘍を切除した、あ

る患者さんの症例です。CTの画像を見ると広範に左右の前頭葉が切除されています。この患者さんは、元は非常に優秀な銀行員だったらしいのですが、術後、出かけるための支度に２時間ぐらいかかるとか、どこで何を食べるか決めるのに、ものすごい時間がかかったり、なかなか決めれなかったりということが起きました。ちょっとした買い物にも、ものすごい時間がかかる、つまり、どれにしようかひたすら迷うようになったのです。一方で、国際問題の議論とかそういうものは普通にできました。知能テストの成績も129点や135点で、99パーセンタイルと非常に優秀な成績を収めています。この知能テストの問題というのは、正解がある問題です。選択肢の中から正解である一つを選んでくださいとか、見本通りに積み木を組み立ててくださいとか。そういう正解のある問題はほぼ100％解けるのに、自分で朝どの服を着るか決めるのに２時間かかるという問題がこの人には起こっています。

昔から「フレーム問題」ということが人工知能の中で言われていて、知識はいくらあっても、こういう正解がない問題というのは、解くべき問題に必要な知識だけを選んできて、それを使って解く必要があるのですが、AIには、それが非常に難しいわけです。例えば、今日どの服を着るかというのも、状況によって考慮すべき条件が違っていて、今日はこういう講演会だから汚いＴシャツじゃダメだなとか思うわけです。だからとりあえずネクタイぐらいしていくか、というふうに我々は思うわけです。天気はどうかとか、シチュエーションをいろいろ考えたりします。でもその時々によって考慮すべき条件が違ったりするので、今日はこの条件でいいけど、明日家でリラックスする時にフォーマルかどうかという条件は気にしなくていいわけです。つまり、考慮すべき条件はその日の状況によって全然違う。その時どきの条件の重み付けも違うし、優先順位も違うし、どの条件は外せるけど、これは外せないみたいなのが毎回違っている。それを我々は経験的に解決しているのですが、そういうものをAIに覚えさせるのは非常に難しいということが昔から言われています。場合によっては条件が両立しないこともある。最終的に絞ったとしても、３つぐらい候補があって最後どれにするかというようなことは、直感みたいなもので決めないといけないのですが、AIにとってはそれがとても難しいのです。

■認知の２つの判断システム

我々は、直感的にいろいろと物事を判断しているという例は、心理学の分野でいろいろ研究されていて、有名な例なので皆さんご存知かもしれませんが、この問題を少し考えてみてください。「おもちゃのバットとボール合わせて1100円で買いましたバットはボールより1000円ほど高かったです。ではボールはいくらでしょう」。次です。「５台の機械を使って５つの製品を作るのに５分かかるとします。100台の機械を使って100個の製品を作るのにいくらかかるでしょうか」。最後、「湖に睡蓮の浮き葉があります。毎日この浮き葉は２倍になります。浮き葉は湖いっぱいになるのに48日かかるとすると、浮き葉が湖の半分を覆うのには何日かかりますか」。皆さんどうでしたか。いろいろ答えを考えていただいたと思うのですが、最初の問題の答えのボールは50円です。おそらく100円と思った方がおられるのではないかと思います。いかがでしょうか。それから次の問題は５分ですが、100分と思った方もおられると思います。次は、これは倍倍になっていくので半分になるのは前日の47日なのですが、24日と思った方もおられるかもしれません。これは簡単な算数の問題のようだけど、多くの人が間違えることが知られています。でも、間違っても心配ありません。いろいろな大学の学生を被験者とした時に、どれぐらい正解したかを見ると、例えば有名なハーバード大学の学生でも、３問全部正解した人は２割しかいなかったのです。全問正解した人の割合を見ても、有名なMITの理系の学生が中心だと思いますが、それでも半分ぐらいしか正解しないという問題です。なので間違っても恥ずかしいことではないのです。

認知の２つの判断システム
(Kahneman, 2011)

System 1	System 2
直感	推論
速い	遅い
自動的	制御的
努力不要	努力必要
連想的	ルール依存
情動的	中立的

実は、人間はどうしてボールを100円って答えてしまうのかというところに、人間の判断の特性があります。AIは、システム２と呼ばれる推論の機能だけを持っています。つまり、いろいろな状況を推論して、ルールに基づいて正確な答えを出すというシステムしか持っていません。しかし、人間はシステム１と呼ばれるような直感的で自動的で情緒的、連想的という性質を持つ早い判断システムを持っています。これは、ノーベル経済学賞を取った研究者のダニエル・カーネマンという、行動経済学を提唱している研究者が最初に言い始めたことです。要するに、人間には、システム１、システム２という２つの判断システムがあって、まずシステム１が直感的に100円じゃないかという答えを出す。それでなんとなくおかしいなと思って、方程式かなんかを立てて解いた人だと、差が1000円で、合わせて1100円だから、バットは1050円でボールは50円が正しいんじゃないかとなって、50円という答えに辿り着いたと思います。つまり、そういう人はシステム２まで働かせたというわけです。ただ世の中には、システム２までいっても答えが正しく出るかどうかがわからない問題がたくさんあるので、ほぼシステム１で解決しているのではないかというのがカールマンの考えです。おそらく日頃の重要でない意思決定は、システム１だけで解決しているというわけです。これはヒューリスティックスと呼ばれていて、短時間に答えを出すための、人間独特の思考方法の１つです。我々の生活においては、最適な答えを出すということは重要なこともあるけど、むしろ適当でもいいけど、短時間でとりあえず間違いじゃないレベルで解答しておけば済むということが実はたくさんあります。バットと合わせて1100円で買ったのだったらボールが50円だろうが100円だろうがそんなに日常生活に影響がないので、そこまで真剣に考える必要はないかもしれません。そういうのは近似解とか局所解といいますが、大体50円か100円ぐらいって思っておけばいい、ぐらいのレベルのことって世の中にたくさんあって、我々はそういうレベルで問題を解いているといえます。AIはそういうレベルで問題を解くということを元々学習していないので、人間とは違う判断にどうしてもなってしまうというわけです。いちいちそういうことを考えていたら日常生活がとても大変だよ、というふうに思われるケースがたくさんあり、人間はそれに対処できるように認知システムを進化させてきたのかもしれません。

ちょっと古いのですが、社会心理の研究ではすごく有名な研究を紹介します。コピー機の前に列ができている時に割り込ませてくれるかという実験です。今だと、銀行のATMのような例が良いのかもしれませんが、昔、我々が学生の頃に、試験の前になると図書館のコピー機の前に、友達からノートを借りて、それをコピーする人たちの長蛇の列ができていました。実際にそういう列ができているところに行って、どういうお願いの仕方をしたら割り込ませてくれるかを試したという研究です。単にお願いのみで、「すいません。５ページのコピーなんだけど、ゼロックス機を使わせてくれませんか？」という例と、無意味な情報をつけて、「すいません。５ページなんだけどコピーをしなくてはいけないので、ゼロックス機を使わせてくれませんか？」と言う、コピーしないといけないのは当たり前なのですが、そういうちょっとだけ無意味な情報で言い訳してみる場合と、さらに「すいません。５ページなんだけど、急いでいるので、ゼロックス機を使わせてくれませんか？」とお願いする場合を設けます。つまり、ただ単にお願いするのと、無味な理由付きのお願いと、意味ある理由付きお願いで、どれだけの人が実際に、割り込んでもいいよって言ってくれるか調べてみると、まず、ページ数が多いとだめと言われる確率は高くなるのですが、ページ数が少ない

時は、単にお願いだけだと60%ぐらいが許可してくれるけれども、なにか無意味なものでも言い訳がついているだけで30％ぐらい割り込ませてくれる割合が高くなって、それは急いでいるのでというのと同じぐらいの割合になるという結果です。割り込ませてくれる側の判断はどうなのかというと、単に割り込ませてっていうと嫌だと言ってしまいがちですが、「コピーをしないといけないので」という、よく考えると何か分からないけど、何らかの理由を言われると、「いいよ、いいよ。」という人が増えるというわけです。これを判断する側は、先程のシステム１のようなものを使って判断していると思われます。自分の前に人がひとり割り込むかどうかという、そんなに自分の生活上大きなダメージがあるわけでもないような状況というのは、結局、そこでの咄嗟の判断で良いか悪いかを決めるのですが、ちょっとでも言い訳がついていると良い方に人間の判断って偏ってしまうということをこの結果は示しています。「今日どこか行かない？」って友達に誘われたら、「良いよ」って言うか、「ちょっと今日は」って言うか紙一重なケースはいろいろあって。そんな時、なんか言い訳があると、「あ、そうか。」となったりすることがあるっていうことがいくらでもあります。そういう判断は、AIがいろんな状況を厳密に精査して下すような判断とは全然違うわけです。ですから、Ａさんはどうして今日は食事に付き合ってくれたのだろうか、ということをAIが理解するのはなかなかに難しいということがこういうことからも分かると思います。

■人間の意思決定の特徴

人間の意思決定の特徴

- 限定合理性と満足化原理（Simon, 1979）
- 環境の複雑さに対する人間の情報処理能力の限界から、完全に合理的な意思決定をすることができない（限定合理性）。
- 意思決定に際し、ある基準を満足した解が見つかれば、それを選択する（満足化原理）。

- 個人差：satisfier vs maxmizer

サイモンという人が言うには、人間は複雑な状況において、基本的には完全に合理的に解決するのは無理だと分かっている、これを限定合理性と言います。また、ある状況で得られる情報だけを使って合理的に物事を100％判断することは無理なので、結局ある基準を満たした答えがあったら、とりあえずそれを選択するという満足化原理というのを人間は使っているのではないか、とも言っています。着て行く服も、究極的には考えてもどれか１着に決まらない状況で、まあまあ今日はこれを満たしてるからこれで行くかというふうに我々は判断します。ただし、どれぐらいのレベルで満足するかというのには個人差があるとも言われています。個人差は領域によっても違う。こだわりがあるところだと、とことん最大限を追求するという人もいるでしょうし、そうでない部分に関しては適当にこんなんで良いやとすごくレベルが低く満足する人もいる。それも人によって違うので、どうしてこの人はこんなレベルで満足しているのかということをAIが理解するのはなかなか難しいと思います。

ここまでのまとめとしては、我々の周りには正解がない問題とか正解でなくて良い問題がたくさんあるということです。結局、こういう状況でAIがタイムリーに適切な答えを出すのは非常に難しい。例えば、割り込ませてくれますかという状況で、AIだったらばいろいろな条件をひたすら考えて、こうなったらこうなって、こうなったらこうなってと言って、答えを出すのにものすごく時間がかかって、結局ダメって言うかもしれないわけです。でも、人間ではそういうことに対して当意即妙に、別に考えずに解答できます。

自動運転とかそういう技術でも、AIは使われていますが、現状はどんなことが起こるかを考えてみましょう。先に述べたフレーム問題が解決されていないので、例えば、自動運転の場面で何かが飛び出してくるとかということは想定しているけれど、上から飛行機が落ちてくるとか、あるいは突然道路が崩れて大きな穴が開くとかといった、AIの想定でない状況が起きた時に、今のところの技術では、人間に「運転を代わって」という指示を出すことになっています。

ですので、フレーム問題は、自動運転の技術ですら解決できなくて、予想外のことが起きた時には人間に代わるという解決を、今のところとるしかないわけです。人間がほんと曖昧な世界でどう意思決定しているかを、100％ AIが理解できる日が来ないと完全な自動運転はなかなか難しいといわれています。

■人は他者の心をどのように理解しているのか

次に、我々は他人の心をどうやって理解しているのかということを考えてみましょう。人間は社会的な生物なので、他の人がどう思っているのか絶えず気にしながら生活しています。特に看護職の皆さんは日頃から患者さんの気持ちとか、何をしたいのかといった、意図とかを絶えず推測しながらお仕事されていると思います。「この仕事お願いします。」と誰かにお願いするにしても快く引き受けてくれる人にお願いするときと、「絶対嫌だ。なんで私にそんなことお願いするの？」と怒り出す人とか、最終的に引き受けてくれるのだけれど、嫌味たらたらの人とかいろんな人が世の中にはいるわけです。そういう人たちに対してどうお願いするのかというのも含めて、我々はこの人がどういう人で、どんな気持ちでいて、どうお願いをしたらどういう行動をとるのか、というようなことに関する膨大な知識を我々は持っているわけです。つまり、他の人の心を察する能力というのを我々は持っていて、AIが鉄腕アトムやドラえもんのように我々の世界に浸透するとなると、最低限そういう能力は持っていないといけないと思います。

そういう能力はいつ頃から獲得するのかを調べる、サリーとアン課題と呼ばれている発達心理学の課題があります。次のような漫画を子供に見てもらいます。まず、最初の場面では、サリーさんはカゴを持っています。アンさんは箱を持っています。次の場面で、サリーさんはビー玉を持っていて、ビー玉を自分のカゴに入れました。さらに次の場面では、サリーさんはどこか外に行ってしまいました。その次の場面で、アンさんは、サリーさんのビー玉をカゴから取り出して自分の箱の方に入れました。最後の場面では、サリーさんが帰ってきました。サリーさんは自分のビー玉で遊ぼうと思いました。そこで、サリーさんは、カゴと箱のどっち側を最初に探すでしょうという問題です。我々は、サリーさんの気持ちになれるので、サリーさんとしてはアンさんがビー玉を移したってことを知らないはずだから、カゴの方を先に探すだろうと普通に思うわけです。4歳から7歳ぐらいまでの子は、サリーさんの気持ちになってカゴの方を探すと答えられるのですが、それより前の子どもだと、サリーさんの気持ちがわかってないので、ビー玉は箱の中にあるという自分の知識をもとに、サリーさんも箱の中を探すと答えます。

心の理論

- サリー＆アン課題（誤信念課題）
- 4から7歳ぐらいで正解できるようになる。

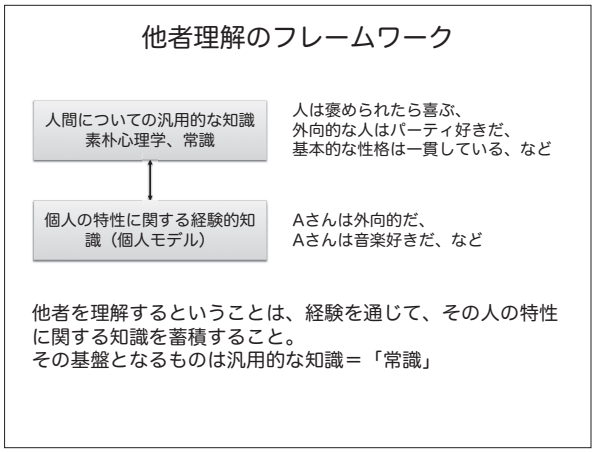
- 正解できるためには、
  - 他人には自分と同じように心があることを理解（like-me）
  - 自分の心の状態と他人の心の状態を区別した理解（different-from-me）
  - それに基づいて他人の行動を予測す（他者の心の理解に基づく行動予測）
  - 共感、信頼の基盤

つまり、この課題に正解するために、サリーさんの気持ちやアンさんの気持ちも含めて、他の人に誰にでも自分と同じような心があるということを理解する必要があります。これをlike-meシステムと言います。そこから先は、自分の心の中身と他の人の心の中身は必ずしも一緒じゃないということをさらに理解し、相手はもしかして自分とは違うことを思っているかもしれないということ、これをdifferent-from-meと言いますが、それが必要です。そういうことができるようになるのが4歳から7歳ぐらいと言われています。つまり、それぐらいになると人の心を理解して、その人がどういう気持ちかというのが分かるようになってくるのです。このような仕組みに基づいて、我々は他の人の意図だとか行動の予測だとかということをしています。そうすることによって、他の人の気持ちに寄り添うみたいなことができるようになってくると、最終的には共感とか信頼と呼ばれるような人間

の社会生活の基盤みたいなものが出来上がります。一方で、自閉症スペクトラムと言われるような人たちは、比較的他者の気持ちを理解したりということが苦手だという報告があったりします。いずれにしても、我々は発達段階で他者の気持ちを理解する能力を身につけていきます。突然こういうところで大声を出しているのは、何かに怒っているに違いない、怒るには何か原因があるに違いないというような、世の中の仕組みと人の心の動き方の仕組みみたいなものを、我々は独自に共通に身につけていきます。

文化に普遍的なものもあれば、文化依存のところもあるのですが、いずれにしろそういう膨大な知識を、我々は身につけているというわけです。例えばロボットを動かそうとすると、我々が持っている知識、常識みたいなものを全部書き出して、全部AIに入れる必要があるということを考えている人たちがいます。MITのOpen Mind Common Senseプロジェクトというもので、世の中にある常識をひたすら集められようとしています。今、100万件以上の常識が集められていて、寒い時はコートを着るとか、太陽はとても暑いとか、晩ご飯の最後にすることは皿を洗うこととか、友人と過ごすことは幸福感を与えるとか、自動車事故に遭遇すると怒りを生むとか、人々に尊敬されたいと思っているとか、人々は良いコーヒーを欲しがるとか。言われてみれば、それはそうだと思うことは我々の周りにはものすごくあるのですが、我々はそういうものを全部知っているから常識的に振る舞えるし、人の気持ちがわかるし、人がなんでこんなときこうしたんだろうか、きっとこうに違いないって推測も成り立つわけです。AIが人と同じように振る舞おうとすると、こういう知識を全部持っていないといけないのかということになってきます。実は、この中に含まれていないことが、まだまだ膨大にあるそうです。なので、常識を全部書き出してAIに教えるのが現実的かという議論もあります。

■他者理解のフレームワーク



結局、我々が他人をどう理解しているのかというと、基本的には一般的で共通の知識、つまり常識を皆が共有していることが背景にあります。人はそもそも褒められたら喜ぶとか、外交的な人はパーティーが好きだよねとか、基本的な性格は一貫しているというようなことは、皆さんが共通に知っています。それと、Ａさんはどういう人かという個別の知識があって、「Ａさん外交的だ。外交的だったらＡさんはパーティーに誘ったらきっと来てくれるに違いない。」と思ったりするわけです。Ａさんについての個別の知識と、一般的に人間はどういうものかという知識の両方を我々は持っています。他者を理解するということは、経験を通じてその人の個人のモデルみたいなものを我々の中に作っていくことです。初めて会った時には、その人がどんな人か分からないけれど、少しずつ付き合う段階で、その人がどういう人かという知識を個別に我々は蓄えていくということになります。個人と親しく接するということは、その人に対する個人モデルみたいなものを我々の中に作り上げていくことで、そのベースになっているものはこの常識と呼ばれるようなものです。だからAIが他者を理解しようとしても、実現には結構時間がかかりそうだということが分かると思います。

我々人間はすごい能力があって、一瞬その人を見ただけでその人がどんな人かってほぼ正しく推測できるという研究がたくさんあります。場合によったら２秒ぐらい動画を見るとか、学会とか講演会でもその人が入ってきて、何か喋

り始めた瞬間にこの人の講演が、面白そうだなとか、つまらなさそうだなってすぐわかったりします。どんなパーソナリティの人かとか、信頼できる人かとか、能力が高い人かどうかとか、利己的そうな人かどうかとか。あと社会経済的状況ですね、金持ちそうかとか。どんな好みを持っている人かというようなことを、ほんの一瞬見ただけでも分かるという研究もあります。例えば、ある人の顔写真をみて、その人がハンバーガーを好きそうかっていうことを答えさせると、二者択一なのでランダムに答えても正解率は50％ですが、実際には、50％よりも高い確率で判定ができます。その時の脳を調べると、前頭葉の内側部分あたりが活動していることがわかります。この辺がthin sliceといって、一瞬の情報だけで人間は判定するのに働いている脳の場所だといわれています。

■他者の理解と行動の予測

結局、我々は人の行動とか顔とかから、その人がどんな心の状態にあって、さらにそれはもともとどんな性格の人であるのか推定しているわけです。それぞれもそうだし、一般的にこういう性格の人はこんな気持ちになりやすくて、その結果こんな行動をしやすいという関係も学習している。我々は他人を判断するときに、こういう構造をそれぞれ常識として我々の中に持っています。ですから、AIに、ある人がどういう人かを理解させようとすると、それぞれ人間はどんな性質の組み合わせからなっていて、そういう性質を持っている人はどんな心の状態になりやすく、結果どんな行動をしやすいのかというような一般常識みたいなものを持たなければなりません。そうでないと、他人というのをAIは理解できそうにないということになります。

AIによる人間の状態、特性の推定

- 行動を指標とした状態の推定
  - 眼球運動
  - 動作、仕草
- 行動を指標とした特性（パーソナリティ）の推定
  - 発話やテキストからのパーソナリティ推定
  - SNSでの選好からのパーソナリティ推定
  - 眼球運動からのパーソナリティ推定

行動からその人はどんな人かということについて、いろいろな解析ができるようになってきています。例えば、視線の動きとか動作、仕草とか、あるいは話した内容とかからでも、その人がどんな人かということが、そこそこの精度で推定できるようになってきています。

例えば、ある人がTwitterとかFacebookとかで発信した内容と、その人の性格の関係を調べた人たちがいます。Facebookに投稿している人たちがどんな言葉を投稿しているかということを集めるのと同時に、その人たちそれぞれにお願いして、パーソナリティのテストをしてもらいます。そうすると外向性が高い人はどのような言葉を使いやすいかとか、神経症傾向が高い人はどのような言葉を使いやすいのかということが分かってきます。逆に、こんな言葉を使う人は神経症傾向が高い人だというようなことを推定することができるようになります。例えば外向性が高い人は、party, boys, girls, beach, weekendとかを呟いているのに対して、内向的な人はanime, internet, pokemon, manga, computerとかを呟いています。こういう言葉を使う人は内向的な人だと、我々は直感的に分かるってことは、我々の知識と大体、一致しているともいえます。神経症傾向が高い人は、lonely, depression, I hateとかそんなことをたくさん呟いていますし、情緒の安定性が高い人はsuccess, basketball, beautiful, beach, soccerとかをたくさん呟いています。我々は、ある人の喋る内容からどんな性格の人かなということが分かったりするのですが、そういうことが実際にAIでも、ビッグデータを解析することで可能になってきてい

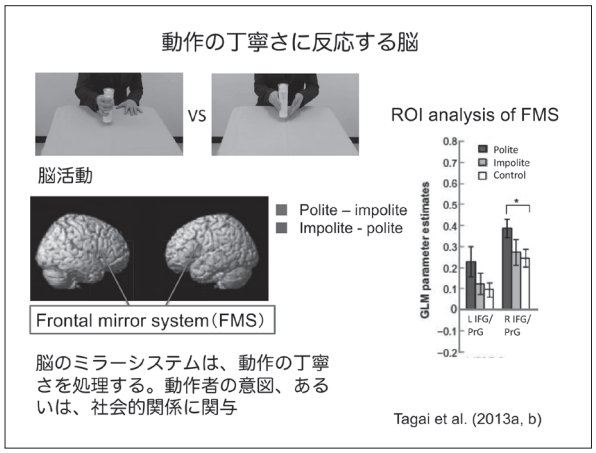
ます。

写真共有サイトで「いいね」の付け方というのも、その人のパーソナリティに依存しているということもわかってきていています。つまり、どんな写真を「いいね」って思うかによって、その人の性格がわかるかもしれないということです。神経症傾向が低い人、つまり情緒的に安定している人は、カラフルな写真や花とか人とかが写っている写真に「いいね」を付けるし、神経症傾向が高い人、つまりとても神経質な人は、比較的モノトーンな写真に「いいね」を付けるとか。あるいは外向性が高い人だと人のポートレートみたいなものに「いいね」をつけます。外向性が低い人は、動物とか風景とかに「いいね」を付けがちです。これは、我々の直感とも一致すると思います。その人の行動からその人の性格のようなものが、ビッグデータを使うことによって、判定できるようになってきているというのが現状です。我々が、こういう行動を取る人はこういう性格だよね、という判断をしているのに用いている知識は、結構集められるようにはなってきているわけですが、我々はものすごくいろいろな情報を使ってこういうことをしているので、その全てを明らかにするのはなかなか難しそうです。

■人間が処理する社会的情報は極めて多様

結局、我々は、SNSで「いいね」を押すかどうかもさることながら、実際には、膨大な社会的情報を使ってコミュニケーションをしています。我々が普段使っていそうな社会的な情報はたくさんあって、外見、動作、顔表情とか顔の形とかもそうだし、言語情報の中でも抑揚とか、「うーん」とかそういう音声、笑い声、泣き声といった非言語的な情報も使うし、対人距離とか位置関係とか、どんな状況で人と会うかというようなものも社会的なメッセージとして相手に伝わります。こういうものを総合して、相手との距離感だったり、相手との人間関係みたいなものを微妙に調節しているわけです。

目は口ほどに物を言うという諺がありますが、世界各地で視線というのは相手とのコミュニケーションに結構重要な役割を担っていると言われています。つまり、我々はアイコンタクトによって相手との意思疎通をしています。発達段階でも、子どもは親の目を見たり、そこから目を逸らすことは生後間もないころからするということが分かっています。そのうちjoint attentionといって、親がパッと目を動かすとそれにつられて子供もそっちに目を動かすということが起き、親が関心を持っていることに子供も関心を持ち始めるということが９から11ヶ月齢ぐらいで起きます。そのうち、逆に、目を道具として使って、欲しいものがあったら、じっと見つめて、親にそれ取ってとお願いするようなことを子供はするようになってきます。目を動かすだけで人はこう動くんだ、目を動かすことによって相手にメッセージを伝えることができるんだということを発達段階で我々は学んでいるのです。ロボットにそういうことまでさせるようにできるのは、まだまだ先の話です。実は霊長類の中で白眼があるのは人間だけで、それ以外にはないので、目を使って何か相手への意思疎通をしたり、意思をコントロールしたりってできるのは人だけだっていうふうに言われています。



あと我々の研究だと、単に人に物を渡すという動作でも、渡し方によって人間の脳の反応が異なるということが明らかになっています。ミラーシステムというのが人の脳の前頭葉にあるのですが、その活動は、単にぞんざいな方法で物を渡されるのと、丁寧な方法で渡されるのとで活動が違うのです。ただ単に物を「はい」って渡すだけでも我々はそこにいろいろなメッセージを込めることができ、それが相手に伝わります。例えば、丁寧に渡されると、結構大事

な物ということや、私に対してこの人はそれなりに敬意を持っているということが伝わります。こういう非言語的な情報も人間の脳の中で処理されていて、そういうものからもメッセージを受け取るわけです。単にものを渡すだけということであっても、いろんなメッセージがそこには込められていることがわかります。

オンラインですと、Zoom疲れ（Zoom-fatigue）が起きるということが言われます。なぜ、オンラインの会話だと対面に比べて疲れるのかということに関する研究が、最近増えてきています。その中では、非言語的な情報が伝わりにくいとか、言語情報がうまく伝わらないことが疲れることの原因だと言われています。我々は、普段何気なく他の人と会話をしているわけですが、それがZoomだとぎこちなくなるということがよくあります。普通に対話している中で、どういうメッセージが対話の中に込められているのかというのは、逆にZoomがなんでぎこちないのかということを知ることで分かってくることがあります。

ある実験の話をします。この実験では、6分間の映像を被験者に見てもらいます。3人の女子学生が会話をしているもので、被験者には、そのうちの1人、リンダと呼ばれる学生になったつもりでこれを見てくださいとお願いします。4分ぐらい経ったところで、リンダがちょっと性的で微妙な発言をし、その後2分間会話が続いてこのムービーは終了します。3つの条件があって、通常条件は普通に自然な会話が流れて、リンダが微妙なことを言っても、それはスルーされて会話が続きます。もう一つはリンダが微妙な発言をした後で4秒間の沈黙を置いて会話が続くようにビデオを編集します。実験後に、リンダが微妙な発言をした後にどれぐらい時間が空いたと思いますかと、単に被験者に時間を聞くと、通常の条件と、中断した条件とでは、その差はほとんどありませんでした。実験に参加した人は、ムービーにちょっとした間があったっていうことは全く気が付いていないわけです。しかし、その時にリンダはどんな気分になったと思いますかというのを被験者に聞くと、中断した時には拒否感とか、負の情動とかが高まったとか、正の感情、帰属感とか同一感なんかが下がったというふうに答えます。つまり、被験者が気づかないような、ちょっとした間があるだけでも、リンダの気持ちとしてはネガティブな気分になるような感じを答えるとうわけです。なので、Zoomとかで対話していて、通信の遅延とかでちょっとした間が空くと、なんとなく無意識のうちにネガティブな気持ちになっているのかもしれません。うまく意思疎通できてない気がする一つの原因は、時間的な遅れがあるからかもしれないといえます。また、すぐに回答しない人は、パーソナリティが低く評価されるという研究もあります。うまく言葉のキャッチボールができない、つまり、ちょっとだけ応答が遅延したりするだけでも、その人の外向性とか誠実性が低いと評価されるそうです。というわけで、我々が気付かないようなちょっとした遅延でも、お互いの一体感が低下していくので、オンラインのコミュニケーションで疲れたり、達成感が得られなかったりする原因はそういうところにあるんだろうと思います。

我々は日常対話の中で、相手とうまく対話のキャッチボールをするために、「うーん」とか「えーと」とかを挟みながら、タイミングを調節しています。あるいは、相手が、ふっと息を吸うところを見て、何か言おうとしてるいというのを感じて言うことを止めたり、というような無意識のコミュニケーションのツールを使って円滑に対話が成り立つようにしています。実は、そういうものがないとロボットとの信頼感などが高まらないと言っている研究者たちもいて、そのために、ロボットにも「うーん」とかそんな感じのことをわざと言わせたりして、自分が今からしゃべるから、ちょっと黙っていてねという情報を相手に伝えることをしたりすることが行われています。つまり人間がするように相手と対話して、相手に不信感とかそういうものを抱かせないようにしようとしています。ロボットとの信頼関係を築くためには、そういうレベルからコミュニケーションを設計していく必要があると言うことです。我々人間がやっていることはかなり精巧にできていて、なかなかロボットでの実現が難しいということですね。

結局、そういう人間が普段用いている社会的なシグナルが多様であるということを考えると、

そういう人間と円滑にコミニケーションするようなAIを実現するのはかなり難しいと言えます。人間は、表情とか仕草など、いろんな方法で情報を発しているので、そのうちの重要なシグナルを見落とさないことが我々の生活にはとても大事です。例えば、「○○先生が怒っているかもしれない」というようなシグナルを見落とさないというのはとても重要だったりするのですが、それを見分ける方法を実現するのはAIではなかなか難しい。また、相手を不安にさせるようなシグナルを発しないということも重要ですが、それをAIが見つけるようにすることもとても大変です。何か聞いた時にほんの一瞬口籠るだけでも、相手を不安にさせる場合が我々の周りにはあります。特に医療場面だとドクターになんか質問した時に、一瞬何か思いとどまっただけで、「何か悪い兆候でもあるのかしら。」と患者さんが思ったりすることがあるかもしれないわけですね。そういうちょっとした間みたいなものも繊細なやり取りの中では効いてくるとすると、やっぱりAIが実現しないといけないことのハードルはずいぶん高そうだということがわかります。

■AIが人間と協調するための最低条件

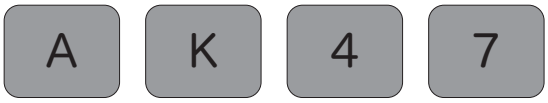
AIが人間と協調するための最低条件

- 信頼
  - 透明性（theory of machine mind, Shariff et al., 2017）
- 納得（交渉、説得）
  - 人間の判断の性質の理解
  - 他者の心の（内容の）理解、感情の理解
  - たとえ話、洞察
- 価値の共有

最後の話題は、AIが人と協調していくにはどうしたらいいかということです。一番は、信頼性が大事で、そのためには透明性が必要だと言われています。つまりAIが何を考えてるのか、何をしようとしているのかというようなことが分かりやすいと比較的不信感が払拭されます。もう１つは、相手を納得させたり説得したりということが、実はAIにはとても重要です。人間はそれが非常に得意ですが、AIにはなかなかそれができません。特に、例え話をしたり、理解したりというようなことがAIは苦手です。

ウェイソンの４枚カード問題

- 設定：４枚のカードがある。片面には数字、もう片面にはアルファベットが書いてある。
- 規則：カードの片面が母音ならば、その裏面には偶数を印刷する。

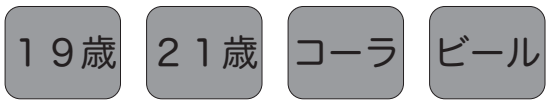


問題：規則どおり書かれているかを確かめるには、どのカードをめくればよいか？（不必要にめくらないこと。）

ここで頭の体操をしたいと思います。ウェイソンの４枚カード問題というもので、AIの分野で昔から言われている有名な問題です。今ここに４枚のカードがあります、片面には必ず数字が書いてあって裏には必ずアルファベットが書いてあります。アルファベットの片面が母音ならばその裏面には偶数を印刷するという規則でこれらのカードが作られているはずですが、ほんとにその規則通りに作られているか確かめたいというわけです。その時、どのカードを調べればこの規則通りに４枚のカードが作られているかを確かめることができるでしょうかという問題です。非常にめんどうくさい課題です。「Ａ」のカードは、母音なのでこの裏は偶数じゃなきゃいけないので、この裏が奇数だったら規則通りではないことになります。だから、まずこれをめくってみようっていうのは正しい判断です。では、それ以外にどれを確認する必要があるかという問題です。実際、有名な大学で実験するとなかなか正解が得られないことが報告されています。まず「Ａ」だけと思った人は間違いです。それと「Ａ」と「４」じゃないかって思った人も間違いです。なかなか分かりにくいですよね。本当は「Ａ」と「７」を調べなければならないのですが、そう言われても何でという感じがしますよね。これはAIなら何の問題もなく解けますが、人間にはとっても解きにくい問題です。

４枚カード問題の変更版

- 設定：４人について片面には、年齢もう片面には飲んでいる飲み物が書いてあるカードがある。
- 規則：カードの片面が２０歳未満ならば、裏面はソフトドリンクでなくてはならない。



問題：規則通り書かれているか（＝法律が守られているか）を確かめるには、どのカードをめくればよいか？

次に、これをちょっと変えるとものすごく簡単になるってことが知られています。最近４人で飲みにいくってことはなかなか難しいですけれど、４人でどこかのお店に飲みに行っているとします。今、４人に、それぞれ自分の年齢と、今自分が飲んでいる飲み物を書いてくださいと言って書いてもらったのがこのカードです。カードの片面の年齢が未成年、つまり20歳未満ならば、裏面の飲み物はソフトドリンクでないといけないということを調べたいとします。つまり、未成年はアルコールを飲んではいけないという法律が世の中にあるので、全員が法律を守っているかを知りたいというわけです。そうすると途端に簡単に答えることができます。ほんとに法律を守っているか知りたければこの「19歳」という人がアルコール飲んでないかってことを調べることと、このビール飲んでいる人が未成年じゃないかを調べなきゃいけないというわけですね。つまり、「19歳」と回答しているカードと「ビール」のカードを調べれば、法律通りになっているかが分かります。実は、これは先程の問題と全く同じ構造ですが、これだと分かったという人が大半ではないかと思います。つまり、我々にとって形式的で抽象的で難しい問題も、何か例え話になるととっても納得できます。我々の知っている日常的な文脈の中でそれが語られると、我々にはよく分かるということになります。このような適切な例え話をAIが考えてくれると、我々は納得できると思います。人間が一番得意なのは、難しい問題を例えで理解することだと思われますので、それに合うようにAIにも理解してほしいというわけですね。実は、比喩みたいなのはAIが苦手としている課題の一つと言われています。一方、人間には、むしろ比喩のようなものを理解する能力があるので、AIが人間を説得するためには比喩みたいなものを上手く使わないといけないという人たちもいます。

■AIを受け入れる側の人間の特性

ここからは、AIを受け入れる側の人間の特性というのも少し考えていこうと思います。人間の側もAIを受け入れる素地というのを持っています。例えば擬人化です。子どもなどは、鉛筆やペンを人形に見立てて、遊んだりするわけです。人間は人間以外のものでも人間らしく見立てることをします。例えば、猫が何か寂しがっているように見立てて、人間の心理を投影するなどです。ですから、AIもある程度人間に近づいてくると、我々の側からAIの側に歩み寄る可能性がありそうだということが分かっています。

我々の実験を紹介します。赤い三角形がぴょこぴょこぴょこと画面の上を動作していくのですが、それを見ただけでも、何か外交的でせっかちぽいなとか、何かモジモジしていて人見知りっぽいなとかという性格のようなものを何となく感じることができます。それをメディアの等式仮説という人もいます。それに関連した研究を紹介します。この研究では、コンピュータを実際に使ってもらいます。コンピュータを使った後で、被験者にはこのコンピュータの性能を７段階で評価してくださいというお願いをします。そして、今使い終わったのと同じPCで、そのPCの評価してもらうのと、同じ製品だけど使ったのとは違うPCで評価してもらうのと、紙と鉛筆で評価してもらうという３段階を用意します。すると、全く同じコンピュータで評価をしているのに、同じPCを使ってそのPCの評価をしようとするとより高く評価する、つまり、我々は相手がコンピュータであってもコンピュータに気を使って高めに評価をしてしまうということがあるというわけです。単に、コンピュータの性能を評価してと言われても、そのコンピュータに気を遣うというようなことを、我々は無意識のうちにしてしまうのです。

というわけで、我々は動物を見てもパーソナ

リティを見い出したりといった傾向があります。犬好きな人と猫好きな人とでパーソナリティが違っていて、犬好きな人は一般的に言われている犬のパーソナリティに近いパーソナリティをしていて、猫好きな人は一般的に猫のパーソナリティだと言われているようなちょっと神経質で内向的でというパーソナリティの人が多いという傾向があったことが報告されています。なんとなく我々は動物とかだけではなく、生き物でないものでもパーソナリティを見出し、それに対して自分に近いとそこに親和性を感じるような傾向もあるらしいので、AIの技術が発展し、スマートスピーカーやペットロボットみたいなものが少し人間らしく振る舞うようになってくるだけで、我々の側から少し歩み寄ってそれにペット的な性質とかそういうものを見出すようになるのではないかと思えます。

特にコロナ禍で対人交流が少なく、制限されているような状況で、話し相手とか相談相手としての需要というのは大きいのではないかということが言われています。我々もそう思っています。特に例えば、施設とか病院とかでお見舞いに行けないとか、一般的な対人の交流が阻害される状況になった時に、話し相手として、AIは活躍するのではないかと考えています。

■AIをどのように社会に受け入れていくのか

最後、どういう人達がそういうペットロボットのような、人に寄り添い、人に近づいていくようなAIを受け入れてくれそうかということです。これはAIとは直接関係ない、我々の研究を見て考えてみましょう。あなたにとって気が置けない人間、ずっといても飽きないしずっと長くいられるような人ってどんな人ですかというのを聞いた研究です。パートナーと答える人、つまり奥さん、ご主人、恋人などと答える人は、女性だと70代まで一貫して全体の半分ぐらいです。さらに女性だと、母親という人が若い人に多く、年齢にしたがってだんだん減っていきます。それに比べて子供とか息子・娘という人達がだんだん増えていきます。一方、男性は一貫してパートナーという人がだんだん増えてきて、60代から70代になると７割から８割の人がパートナー、つまり奥さんとか恋人というわけです。そうすると奥さんの側から見ると、全然相手にしてないのに、男性はとっても奥さんを頼りにしているという非対称性が見て取れます。もしかしたら、こういうミスマッチのある人達にパートナーロボットみたいなものが受け入れられる可能性があるのではないかと考えたりするわけです。そのようなことで、コロナ禍で様々な人的交流が制限されてくる世の中になった時に、スマートスピーカーが質問に答えてくれるとか、相手にしてくれるとかといった形で、ペットロボットをもう少し人間らしくすることによって、弱いAIが考えるようなとっても狭い領域を解決するとか、専門的な問題を解決するわけではないけど、我々の孤独感とか不安感みたいなものを軽減してくれるような役割が浸透していくのかもしれないと思っています。

皆様の職業に近いところだと、施設などで、顔色や健康状態をモニタリングしながら、入所者を見守っているようなロボットとかがこの先普及していく可能性は大いにあると考えています。ですので、強いAIに関してはなかなか色々な問題があって、にわかに鉄腕アトムやドラえもんみたいなものができるという状況ではないのですが、だんだんそちら側にいろんな研究が進歩していることは確かです。

一方で、この弱いAIは技術が進展してきてるいので、ほんとにピンポイントであれば、かなり高度なことができてくるようになります。そちらの側はそちらの側で、我々の方に受け入れる素地があるのかというのは重要な問題です。例えば自動運転とかそういう車が進展してきたときに、例えばそれに家族が犠牲になっても許せるのかとかですね。AIのセールスマンが何か高いものを説得して売り込もうとしてるときに、納得して買えるのかとかですね。AIが治療方針とか診断を下したときに我々はそれを俄に納得できるのだろうかとかです。例えば、余命がビッグデータから解析できて、あと１ヶ月しか余命がなく、治療しても回復する見込みないから止めときましょうというふうにAIが言った時に、家族や当人としたらそれは納得できるのか、ということを考えたりするような時代がおそらくすぐにやってくると思います。こういうAIは答えを出すのは得意で、ほとんど間違えませんが、

そこを上手く、相手が納得するように伝えるようなことを実現するのは、現状のAIでは程遠いので、おそらくは専門職の中でもそういう役割分担がなされていくものと思います。正確に答えを出すのはAIだけど、それを相手が納得するように伝えて、相手と信頼関係を結びながら最終的にいい方法を探ってくようなことは、人間の専門職でないとできないのかもしれないと思ったりもします。

以上で今日お話ししたいことは終わりですが、結局AIそのものがこれから大幅に進歩していくことは確かで、いろんな場面に浸透してくることも確かです。今のところ、技術的な問題もあって、なかなか鉄腕アトムとかドラえもんみたいなものができるとは思えないけど、今までそんなものは受け入れられないと思っていたところにもAIが浸透していって、人々の不安感とか孤独感を和らげるようなことができるぐらいのところには来ているのではないかというのが技術の現状です。