気になるこの用語

第30回

消費生活相談の周辺用語を取り上げ、やさしく解説します。

松原 仁 Matsubara Hitoshi 人工知能研究者
東京大学大学院情報理工学系研究科AIセンター教授。公立はこだて未来大学特任教授。元人工知能学会会長。著書に『AIに心は宿るのか』(集英社インターナショナル、2018年)など

人工知能 ディープラーニング(2) 画像認識AI　説明可能AI

ディープラーニングは従来の人工知能の機械学習と異なり、ある対象の特徴を自動で学習できるというすごい能力を有しています。一方で、自分の出した答えをどのように導いたかを説明することができません。

ディープラーニングのどこがすごいか

ディープラーニングはそれまでの機械学習の技術および人工知能全般の技術とどこが異なるのでしょうか。一言でいうと、人間が教えなくても対象のポイントとなる特徴が何かを自動的に学習できるところにあります。これをコンピュータに猫の画像を認識させる例を使って説明しましょう。

ディープラーニングが登場する前から従来の人工知能はコンピュータに猫の画像を、それなりの精度で認識させることは可能でした。ただし猫を猫と認識するための特徴は、人間がコンピュータに与えていました。「4つ足で、耳が立っていて、ひげが生えていて……」という特徴を持っているものは猫と認識しなさい、と人間がコンピュータに教えていたのです。犬の画像を認識させるときも、同様に犬の特徴を人間がコンピュータに教えていました。コンピュータは人間から教わったとおりに画像の特徴をチェックしていただけなのです。

こういう説明を受けると次のような疑問を持つと思います。「コンピュータが猫の画像を認識できるというが、コンピュータは人間が教えたままにやっているだけではないか、賢いのはコンピュータではなくて猫の特徴を示した人間のほうではないか」という疑問です。この疑問はもっともで、猫の特徴というデータを与えられたら猫の画像を認識できるのは当然といえば当然です。結果的にコンピュータが猫の画像を認識できるようになったのだからよいではないか、と思われるかもしれませんが、問題があります。猫の特徴はこれこれ、犬の特徴はこれこれと対象ごとに細かく人間がコンピュータに教えてやらなければならず、人間の手間が膨大にかかってしまいます。

ディープラーニングは従来の人工知能の機械学習とは異なり、人間が猫の特徴をコンピュータに教えることはしません。その代わりに猫の画像を大量にデータとしてコンピュータに与えます。同様に猫ではない画像も大量に与えます。その際にどの画像が猫で、どの画像が猫以外かの正解情報も教えます。ディープラーニングはそれらのデータから猫の特徴を自ら獲得できるのです。コンピュータはディープラーニングで学習したのちに、人間が猫の特徴を教えることなく、猫の画像を高精度で認識するようになります。これは人工知能の研究にとって画期的なことでした。従来の人工知能の機械学習では、前述のように賢いのは教えている人間だったのですが、ディープラーニングはコンピュータが賢くなったといえそうです。いちいち人間がコンピュータに対象ごとの特徴を教える必要が無くなりました。ただし、人間が特徴を教える代わりに大量のデータを与える必要があり(猫の特徴をコンピュータが自習するためには数万枚の猫の写真が必要でした)、それはそれで違う手間がかかります。

ディープラーニングのブラックボックス問題

すごい能力を持つディープラーニングですが、問題点もあります。これまで説明したように学習には大量のデータが必要ということがその1つです。別の大きな問題点として、どうしてコンピュータはディープラーニングでその結論に達したかを、人間が分かるように説明できないこと(ブラックボックス)が挙げられます。前回、「アルファ碁」が囲碁の名人に勝ったことを取り上げました。アルファ碁は名人に勝つぐらいなので、とてもいい手を打つのですが、その局面で、なぜその手を選んだのかを囲碁ファンに分かるように説明できません。

一方、プロの囲碁棋士は、なぜ自分がその局面でその手を選んだのかを囲碁の用語を使って説明することができます。あくまでアルファ碁は「いい手を打つだけ」なのです。アルファ碁はディープラーニングで計算して最善手と判断した手を選んでいるので、どのように計算が進んだのかを追えば、確かにその手が最善と判断したプロセスは確認できるはずです。しかしそれは囲碁の用語での説明にはなりません。また、一手を決めるのにとても多くの計算をしているので、人間がその計算を追うのは容易ではありません。

囲碁はゲームなので手の説明ができなくてもいい手を打って勝てればよいかもしれません。しかし例えば、ディープラーニングによる医療診断の場合であればどうでしょうか。人工知能が患者のデータからどの病気かを判定して、飲むべき薬を提案したとします(これは部分的には既に実現していることです)。人間の医者であればデータからどのように診断をして結論に至ったのかを、患者が分かるように説明してくれます。患者はその説明に納得して、処方された薬を飲むわけです。ディープラーニングは診断結果を出すだけで、どのようにしてその結果に至ったのかの診断プロセスを説明してくれません。これでは信頼して命を預けることは難しいといえます。仮にディープラーニングによる診断結果が常に100%正しいなら、説明が無くても診断結果を信じて薬を飲むかもしれません。しかしながら、ディープラーニングはすごい技術ではありますが、絶対に間違えないわけではありません。人間の医者もまれに誤診をするように、ディープラーニングでもときに間違えます。囲碁プログラムも常に最善手を打つわけではありません。ときに間違えます。名人よりも間違いが少ないので名人に勝ちますが、決して間違えない囲碁の神様ではないのです。ときに間違いを犯すディープラーニングが満足な説明ができないのであれば、重要な場面でディープラーニングの結論を信頼することはできません。

結論を得たプロセスを説明できないというのは、人間の意思決定を支援する道具としては致命的な欠陥ともなります。世界中で多くの人工知能研究者がこの問題点の解消に取り組んでいます。説明ができないディープラーニングに対して、説明ができる人工知能のことを「説明可能AI」と呼んでいます。

ところで、ディープラーニングが答えを得たプロセスをうまく言語化して説明すれば人間はそれを理解できるでしょうか。それほど簡単ではありません。囲碁のプロ棋士が手を決めた思考過程をそのまま説明しても、おそらくプロ棋士よりも弱い囲碁ファンは、難し過ぎてその説明が理解できないでしょう。解説がうまいプロ棋士は、囲碁ファンのレベルに合わせて適切な説明を作っていると思われます。それは答えを導いたプロセスそのものではなく、理解しやすい説明用のプロセスなのです。人工知能もそのような理解しやすい説明をすることが期待されています。

ところで、ディープラーニングは、任意に選ばれた対象の特徴を自ら獲得できるところがすごいという話をしました。これはよいことだけでなく、思いがけない問題も引き起こします。次回はその説明から入ることにしましょう。