

# JAICO

産業カウンセリング

# 10

2016.No343

特集 人工知能は  
カウンセリングの夢を見るのか?

ニュース 第2回理事会報告  
臨時総会報告

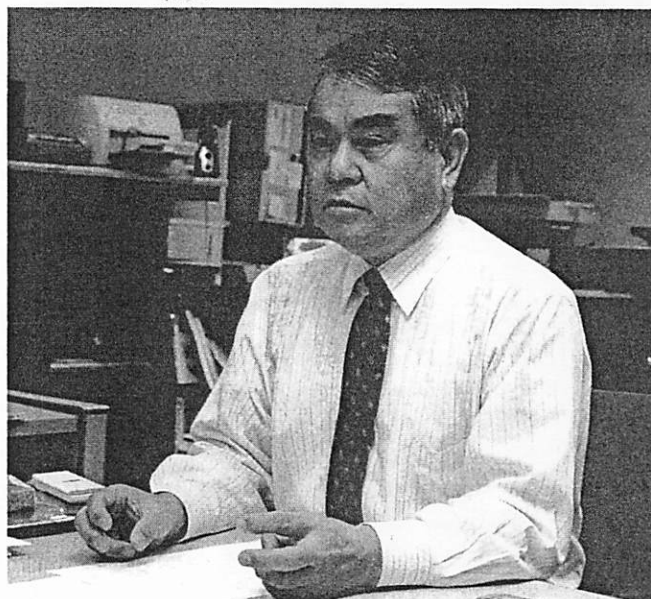




# 人工知能が 人間の気持ち を理解する日

東京工科大学教授

亀田弘之



ペッパー君よりも先に感情を持つロボットを開発。人工知能が知らない単語を獲得するプログラムや認知症の高齢者の訓練に使うゲームソフトなど、多様な研究で注目を集める亀田弘之先生に人工知能の感情についてお話を伺いました。

私が人工知能の感情を研究してみようと思いついたのは、人間と人工知能が本当の意味で対話できるようになるためには何が必要なんだろうと考えたからです。

もともと私は人に興味がありませんでした。数学をやっていたら楽し

いという人間でしたから。ところが、あるとき人間の内面の素晴らしさに気づいたんですね。時間空間を超えた人間存在。2000年、3000年前の文献を読んで理解し、自由にその時代の思想を時空を超えて学べるのが人間なんだと。それはすごいことだと気づき、自分もその人間なんだと感動したし、嬉しかったんですね。そこから人間に興味を持つようになったのです。

じつは人工知能における感情の研究も人間理解を想定したものです。だから感情が変わっていく、それによって行動が変わっていく因果関係モデル

を作ればいかなと考えています。

最近、部分的に脳機能の特性を持つニューラルネットワーク(※1)やディープラーニング(※2)といった人工知能の手法が話題になっています。コンピュータが自力で理解を深める機械学習の活用で、人工知能分野でさまざまな成果を上げています。グーグルの人工知能「アルファ碁」がプロ棋士を負かしたことも話題になりました。しかしニューラルネットワークを使った人工知能は、因果関係がハッキリしないのです。実際、「アルファ碁」開発者も、どうしてこんなに素晴らしい手を人工

知能が考えたのかわからないと発言しているとのこと。つまり入力すれば、何か出力するというニューラルネットワークは、ゲームや経済など目的を持った問いかけに答えを出しますが、人間理解に大きな力を発揮するとは思えないのです。

感情を持つ人工知能開発の私のキーワードは計算論的、構成論的モデルです。コンピュータは何らかの計算モデルに基づいて感情を生み出すと仮定して研究するのが計算論的仮説に基づいてシステムを組み、実際にそのシステムを動かしながら仮説を検証していくのが構成論的なアプローチです。私の言う構成論は、人間の脳もある種のコンピュータであるという仮説に基づいています。脳がコンピュータのOSに相当するとしたら、OS同様に設計できるのではと考えたのです。

人工知能における感情の研究を始めるに当たって、まず感情を表現することから取り組むことにしました。ここで問題となったのが、感情とは何かということでした。これは数年間議論しましたね。

国語辞典にはたくさんの感情表現が載っています。また、脳神経科学系の論文の中には、感情は120余りと書いてあるものもありました。しかし同じような感情が連続的に変化し、それぞれに名前がついていたりもするので、つまり感情の数は切り取り方によって大量にもなるし、逆に感情に関する辞書の単語を全部抜き出してもすべてを網羅したことになるのではないのです。そこで工学的な立場を取ることにしました。アプリケーションの目的を達成できる感情を用意・設定することにしたのです。例えば、カウンセリングなら細かい感情の設定まで必要でしょう。しかし、単におしゃべりして楽しいといったレベルなら感情の種類を絞っても問題ないというわけです。

そこで私は感情を喜怒哀楽だけに限定。そこに「ご機嫌パラメーター」という機能をプラスしました。嬉しい、悲しいだけではなく、機嫌によってロボットの発言や行動が変わるようにプログラムしたのです。

ロボットを動かす人工知能には、「心のモデル」という名での感情モ

デルが設定されています。感情のモデルに応じて、なじられたら悲しくなり、悲しそうに話すよう設定しました。また同じことを言われても、機嫌によって言動も変わります。例えば、「お腹いっぱいのときに「何か食べる？」と言われても嬉しくありませんが、空腹なら嬉しく感じるので、このレベルが第一段階です。

第二段階では、人間の持つ気持ちや意図を推察したり理解したりできるようにしました。仕組みは、人工知能が持つ「心のモデル」とまったく同じものを人間も持っていると考えられるというものです。例えば、「お腹が空いたときに「何か食べる？」と言われたら、人工知能は非常に嬉しい。だから人間に「何か食べる？」と聞いて嬉しそうなら、お腹が空いているのだろうか？と考えるわけです。自分だったらこうだから相手も同じだ、という理解です。異文化コミュニケーションの反対ですね。

このようなプログラムで人工知能が感情を持ったといえるのかという疑問があるかもしれませんが、接する人間がそう感じたのなら感情を

持ったといっていると思います。とはいえカウンセリングで大切な共感が可能かという点、そもそもロボットはぶつかっても痛くありません。つまり痛いふりしかできないという問題もあるのです。



複雑さを整理する

こうした限定的な感情を増やしていけば、人工知能はもつとさまざまな反応を示すはずです。しかし人工知能を人間理解につながるためには、仕組みが透けてみえる、冒頭でお話した因果関係がわかるシステムにする必要があります。

では、人工知能の仕組みを理解するためには、どうしたらいいのでしょうか。そのカギとなるのが、複雑に入り組んだ構造の整理です。

今やスマホですら7000万行、8000万行のプログラムが組まれています。1年前のある高級車は1億行のソフトで動いていました。ここまで複雑だと人間が理解するのも容易ではありません。とはいえ車ですから、プログラムのどこかにバ

グがあれば、人が死んでしまう危険性もあります。

実はソフトウェアが複雑化する以前、ハードウェアがかなり複雑になりました。その解決策がモジュール化と階層化でした。グループ化して、階層に分ければ、個々を理解できなくても、システムとして理解できます。感情についても、同様の形で整理していく必要があると考えています。例えば身体に直結した快・不快の情動を下層に、社会的な感情を上層のレベルに分けておくというイメージです。

ある程度の研究成果は、だいたい10〜15年で発表されるものです。今はスピードが早いので、5〜10年で結果が出るかもしれません。感情を持つ人工知能の研究も、それぐらいの期間で、何らかの成果を出すのではと考えています。

※1 ニューラルネットワーク…脳機能に見られるいくつかの特性を計算上のシミュレーションによって表現することを指した数理ネットワークモデル  
 ※2 ディープラーニング…人間の脳神経回路をモデルにした「機械学習」の手法の一つ。「ニューラルネットワーク」がベース