

明日への
トビラ

Vol. 12

ひとの知能にせまる 画像認識技術

膨大なウェブ情報がもたらす未来

SNS や写真投稿サイトの普及により、ウェブ上には膨大な量のラベル付き画像データが蓄積されている。東京大学の原田達也教授は、これらのデータを活用して、有害情報の検出・削除や近未来のウェアラブル（装着型）情報端末などにも応用できる画像認識技術を研究している。生活空間にあるすべての「もの」や「シーン」を、あまねく認識できるコンピューターが実現すれば、能動的に活動するロボットの誕生も夢ではない。



実験的にスマートフォンのアプリを開発。認識した物の名前を画面に表示する。



大量データで コンピューターを賢くする

「万物を認識できる人工知能をつくりたい」と夢を語る原田さんは、ロボットの研究者だ。ロボットがその眼でとらえた景色を的確に認識できるシステムをつくらうと取り組んできた。

2008年には、その試作品となる「人工知能ゴーグル」を開発した。小型カメラ付きのゴーグルで装着者が見た映像を記録し、それが何であるかという文字情報をモニターに表示するシステムだ。例えば、ゴーグルを通して花を見ると、即座にモニターに「flower」の文字が表示される（p.13右上）。カメラがとらえた情報と、コンピューターが記憶している情報との関係性を計算して、文字情報を表示する。

研究室では、取り込んださまざまな情報間関係性を計算する手法（アルゴリズム）を工夫することで、画像認識の精度を高めている。精度の向上には「計算」が重要だが、画像を認知、理解するためには、「知識も必要です」という。

海外の大手写真投稿サイト「Flickr」などの普及により、ウェブ上にはラベル付きの画像データが膨大にある。コンピューターに知識を持たせるために、それらの画像や文字情報を集めて、関係性を学習させた。アルゴリズムに工夫を重ね、大量のデータから画像の特徴を高速に正確に引き出す技術をつくり上げた。認識スピードと精度の向上により、2012年には、画像認識技術の詳細な物体のカテゴリを識別する部門で世界1位、一般的な物体のカテゴリ識別では2位の成績を収めた。

このシステムは、画像に単語を関連付けるだけに留まらない。例えば「犬」「人」「噛む」という単語がラベル付けされた場合、犬が人を噛んでいるのか、人が犬を噛んでいるのか、文字情報だけではわからない。原田さんは、単語ではなく、意味のある文章で「犬が人を噛んでいる」とラベル付けができるシステムも開発している。



有害情報の検出・削除や 忘れ物支援などに応用

この画像認識システムは、皮膚の症状診断など「目で見て判断」することであれば、何にでも応用できるという。

実用化に近いものとして、ウェブ上の有害情報を排除するフィルタリングがある。現在の画像監視は、大量のデータを人の目で1つ1つ判断して選別している。ショッキングなシーンなど、心理的な負担になる画像もウェブ上にはたくさんある。コンピューターによるフィルタリングができれば、こうした負担を減らすことができるし、作業の省力化にも貢献できる。企業との共同開発も進み、SNSやウェブ上の投稿から不適切な画像を自動検出するサービスを今年7月に始める予定だ。

眼鏡型や腕時計型の情報端末の開発が進み、今年は本格的に市場導入されるとみられている。もし、カメラ付きのウェアラブル端末でとらえた情報が、常時そのままサーバーに送信されるようになれば、そこに写された他人のプライバシーが大きな問題となる。最低限のプライバシーを守るためにも、画像の自動的な選択と排除は欠かせない。

一方、映像への文字情報の付加で可能になるのは、「忘れ物支援」や「動画要約」などである。例えば、眼鏡型端末で動画を撮影しながら文字情報も付加しておく。携帯電話が行方不明になったとき、「携帯電話」をキーワードに検索し、最後に見たときの映像が出てくれば、忘れ物を探す助けになる。

また、動画を言語化することで、これまで不可能だった中身の検索も可能になる。長時間の動画をいちいち早送りで見なくても、文字情報を頼りに内容を確認し、見たい部分の動画を瞬時に頭出しすることができる。「ウェブ上に投稿される動画が急増しているだけに、非常に重要な技術になってくるでしょう」と強調する。

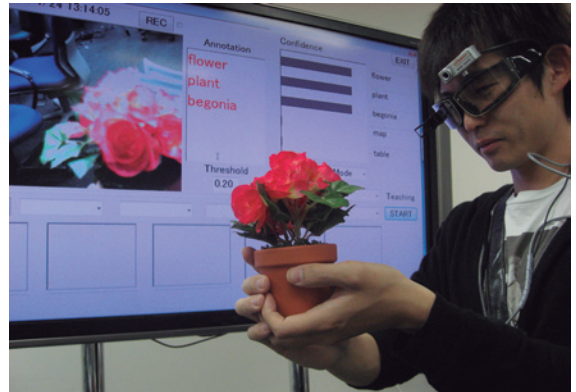


Trees and bushes in a blue sky in the background

画像認識の事例。コンピューターは上の写真を色や形だけでなく情景を認識し、英語で文字表現をした。

画像認識システムの出力例

2008年に開発した人工知能「Google」がカメラが映したものを認識した結果が、後ろの画面に表示される。鉢植えの花を見て導き出された言葉は「flower (花)」「plant (植物)」「begonia (ベゴニア)」(画面中央)。色や形のバリエーションが豊富な花の中から、正しくベゴニアを認識した。将来これが実用化すれば、図鑑データとともに自然観察もできるようになるだろう。



「気づき」機能で実世界を認識する

人間は前後や周囲の状況を見ながら、周りを的確に認識している。それに対してロボットの頭脳となるコンピューターは、画像に映るものの優先順位や価値判断ができない。例えばフルートを持った女性の写真を見て、フルートに目がいくか、女性に目がいくかは、人間の場合は状況によって判断するが、コンピューターにはその判断がつかない。

「画像を認識するときには、どこに着目するかが重要です。それが人間のような知能をつくるための根源なのです」という。

私たちが日常生活で目にする実世界は無限ともいえる。コンピューターは、ウェブ上の情報は認識できても、実世界を認識するのはまだ先のこと。しかし、人間も目に見えるすべてを認識しているわけではなく、一部を切り取って認識している。コンピューターが実世界を認識できるようになれば、人の経験を広げ、これまでできなかったことも実現できる可能性を秘めている。「人間の能力を超えるかどうか」。高いレベルの知能づくりに挑戦しているワクワク感が、この研究の醍醐味だという。

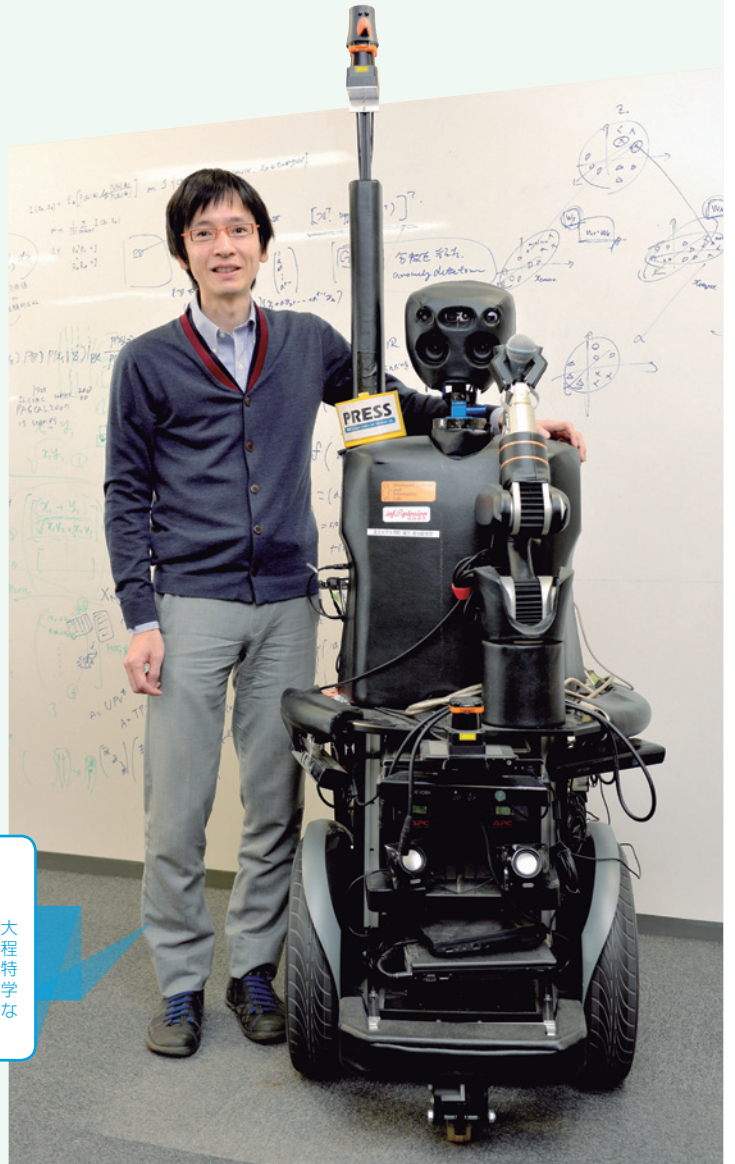
一般的にコンピューターの画像認識の性能は、毎年約10%の割合で上がっている。「さすがに頭打ちだろう」と思っていると、どこかでブレークスルーが起きて、想像以上に進展するのです」。

人が倒れていたら「異常」と判断し、助けを呼びに行く。見たものを瞬時に言語化し理解する。高度な画像認識技術の進展で、人間のような「気づき」機能を持つロボットが身近に登場し、活躍するのも、そんなに遠くはなさそうだ。

原田 達也 はらだ・たつや

東京大学情報理工学系研究科 教授

東京大学工学部卒業。2001年、東京大学大学院工学系研究科機械工学専門課程博士課程修了。博士(工学)。00年、日本学術振興会特別研究員。01年、東京大学大学院情報理工学系研究科助手、06年同講師、09年同准教授などを経て、現職。



マイクとレコーダーを備えて動き回るロボットは、名付けて「ジャーナリストロボット」。「何かおかしい」を感知し、自ら「ニュース」を見つけ出し、「これは何ですか？」と能動的に取材することもできる。実験を繰り返しながら知能を与え、研究室で育てている。研究室の壁には一面に数式があふれている。