

いま「AI+ロボット」が熱い

選考委員 永田 和之



昨今の AI ブームのおかげで、近い将来にロボットは何でもできるようになると思われている。これは、期待を込めて言うと半分正しいが、現実はその甘くは無い。ロボットの最も基本的な作業に、棚や箱の中に置かれた物を掴んで取り出す作業がある。この作業をピッキング作業という。ピッキング作業ができなければ、ロボットは人と会話をしたり、踊ったりするだけの面白い機械でしかない。それはそれで良いが、やはりロボットの主要なアプリケーションは、人が行っている作業の代行である。そもそもロボットの語源は、チェコ語の「労働」を意味する *robota* である。

しかし、こんなことを言うと驚かれるかもしれないが、ロボットは実世界で最も基本的なピッキング作業ができない。でもロボットは、工場の中で部品を掴んで機械を組み立てているのではないかと、言われるかもしれない。現在、工場の中で稼働しているロボットの大部分は、綺麗に並べられた単一種類の部品のピッキング作業を行っているに過ぎない。単一種類の部品のピッキングでも、トレーの中に複数個の部品が乱雑に入れられている場合は、とたんに問題が難しくなる。この問題をビンピッキングと言う。ビンピッキングでは、まず部品の集合の中から個々の部品を分離して認識する必要がある。次に、最も掴みやすい部品を選定し、どこをどのように掴むのかを決定してやる必要がある。ビンピッキングは、ここ数年、ロボット工学の分野で地道に研究され、ようやく実用化の見通しがついてきた。ところが、複数種類の部品が箱の中に乱雑に置かれている場合には、個々の部品が何であるのかを特定する必要がある、問題が更に難しくなる。それでも工場の中では、部品の種類が限定されており、CAD モデルも得られているのでまだましである。ロボットが工場から飛び出して、物流倉庫や店舗、オフィスや家庭内といった我々に近い実世界で活躍しようとした場合、扱う物品の種類は膨大になり、同じ種類の物品でもデザインによって様々な形状であるため、個別の物品を特定することは至難となる。この問題は、機械学習の分野で、一般化物体認識の研究として盛んに取り扱われている。ロボットが物流倉庫や店舗、オフィスや家庭内で活躍するためには、互いに重なり合った多種多様な物品の中からターゲットを探し、周りの物品との接触を考慮しながらターゲットを取り出す、実世界でのピッキング問題を解決しなければならない。現在、「AI+ロボット」の分野で最もホットな話題は、実世界でのピッキング作業に関するものである。

最近、ロボット業界で話題になっているもので、物流最大手の Amazon が主催する、Amazon Robotics Challenge (昨年までは Amazon Picking Challenge) というロボット競技会がある。これは、物流倉庫でのピッキング作業を想定した競技会で、今年で 3 回目となる。今

年は7月に名古屋で開催された。ちなみに昨年はドイツのライプツィヒでの開催。今年は、MIT、カーネギーメロン大学、カールスルーエ工科大学といった錚々たる面々の16チームが参加した。日本からは、MC²（三菱電機・中部大・中京大）、NAIST-Panasonic（奈良先端科学技術大学院大・パナソニック）、Team T2（東芝・鳥取大）、Team K（東大）の4チームが参加した。競技は、Stow（詰め込み）とPick（取り出し）の2種類の作業を実施し、得点を競うものである。Stowはトレイに入っている20個の物品をコンテナに詰め込むもので、制限時間内にできるだけ多くの物品をコンテナに詰め込むことを競う。物品の半分は事前に公開されているが、残りの半分は競技の30分前まで未公開である。未知の物品を詰め込むと得点が高い。Pickは32個の物品の入ったコンテナの中から指定された10個の物品を取り出して3つのトレイに仕分けするというもの。いずれも、複数種類の物品が乱雑に重なり合った難易度の高いピッキング作業である。競技は準備を入れて4日間で、2日目にStowを、3日目にPickを競い、StowとPickの合計点の上位8チームが4日目の決勝戦に臨む。決勝戦はStowとPickを連続して行う。優勝賞金は8万ドルである。今回、日本から決勝戦に進出したのはNAIST-Panasonicのみで、優勝したのはARCV（オーストラリア・ロボティック・ビジョン・センター）であった。ARCVはクイーンズランド工科大学、アデレード大学、オーストラリア国立大学、モナシュ大学の合同チームで、コンピュータビジョンの分野が強いことでの知られている。

さて、各チームのロボットハンドを見ると、エア吸着方式と二本指グリップのハイブリッド型が多い。中にはエア吸着方式のみで勝負しているチームもある。どのチームも、二本指グリップを用いることは稀で、流量の多い強力なバキュームで物品を強引に吸着し、コンテナやトレイに放り込んでいた。二本指グリップよりもエア吸着方式の方が容易に物品把持できるのだ。競技では、丁寧に物品を扱ったり、きれいに物品を詰め込んだりすることは求められていない。要求されているのは、乱雑に置かれた複数種類の物品の中から、できるだけ多くの物品をピッキングすることである。つまり、質より量が求められている。技術的には、カメラ画像を入力とし、どの物品を掴むのか優先順位を付けた後に、その物品の掴み方を出力する機械学習のアルゴリズムを競うものとなっている。この中で、画像認識の学習アルゴリズムの優劣が勝敗を左右しており、コンピュータビジョンが強いARCVが優勝したのもうなずける。

競技を観戦すると、ターゲットを正しく認識できなかったり（このケースが一番多い）、把持した物品を搬送途中で落としたり（ハンド設計ミス or 把持戦略ミス）、他に把持できる物品があるのに難易度の高い物品に何度もトライして無限ループに陥ったり（ターゲッ

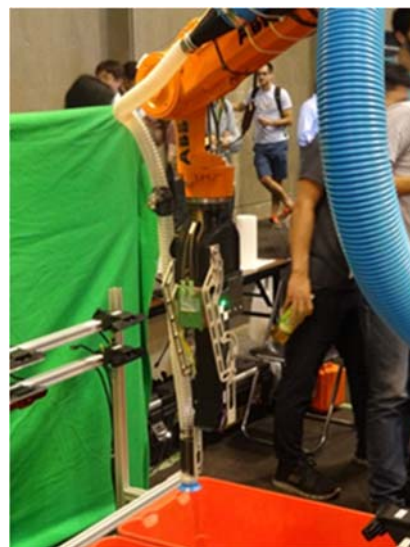


図 Amazon Robotics Challenge
(MIT-Princeton チームのハンド)

トの優先ミス) と、苦戦しているチームが続出していた。今回の競技は、ピッキングの質よりも量が優先され、ターゲットの種類や数も限定されているため、実世界でのピッキング作業と比べて易しい設定である。それでも世界でトップクラスの各チームは悪戦苦闘しており、中には一度もピッキング作業に成功しなかったチームもあった。我々は実世界で、当たり前のように目で見えたものを掴んで取り上げることができる。しかしロボットで同じことをするのは、とてつもなく難しいのだ。

では、ブレイクスルーはあるのだろうか。ペンフィールドのホムンクルスというのがある。これは、大脳の運動野・体性感覚野と身体部位の関係を表した小人である。この小人を見ると、巨大な両手の間に小さな体と顔が乗っている異常な姿が描かれる。これほどに、脳は手のために多くの情報処理を行っている。手の運動に係わる知能は暗黙知と呼ばれ、普段我々の意識にのぼらない知能である。このため、脳は手のために何を計算しているのか良くわかっていない。暗黙の知能を演繹的に表現しコンピュータに実装するのは難しいのである。しかし、最近の計算機科学の進歩により、大規模データが扱えるようになり、ロボットが経験したデータをクラウドデータベースに蓄積できるようになった。また、蓄積した膨大な経験データを用い、機械学習によりロボットは新しい行動を獲得できるようになってきた。最近 Google は、14 台のロボットを用いて 2 ヶ月間にわたり 80 万回ものピッキング作業を繰り返し、人が思いつかないような物品の掴み方を獲得している。クラウドデータベースと機械学習の組合せによる帰納的アプローチは、ロボットに経験を積ませることができる。この帰納的アプローチは、ロボットに大きなブレイクスルーをもたらすような気がしてならない。

(国研) 産業技術総合研究所