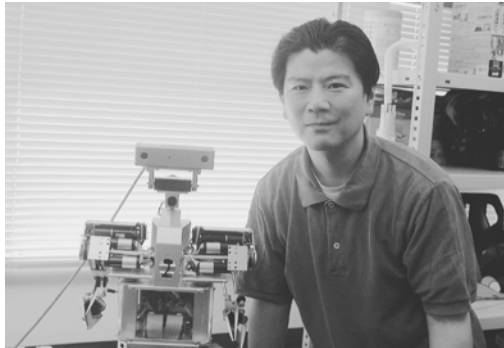




# 人らしいロボットをつくる

## 長谷川 修 研究室～像情報工学研究施設



長谷川 修 助教授

漫画や小説、映画などの様々な物語にはよく人らしいロボットが登場する。彼らは人々に夢や感動を与え、人に憧れを抱かせた。今日ロボットの開発は急速に進められているが21世紀になった今もアトムのような人らしさを持ったロボットは生まれていない。人が当たり前のようにしている動作をロボットにさせる難しさに多くの研究者たちは頭を悩ませてきた。

今回は「学習」という、人と機械では根本的な違いがある行動について研究し、実際にロボットを制作している長谷川研究室を紹介しよう。

### 人になれない機械

今日人のように振舞う「ヒューマノイドロボット」が注目を集め、様々なロボットが開発されている。HONDAのASIMOや、SONYのQRIOなどといった多くの人が見聞きして知っているだろう。まるで人のように歩いたり、握手をしたり、指示されたことを実行したりする様子は、見ている人間に人らしさを感じさせる。しかしこれらのロボットは動作が人らしくみえるだけで、本質的な部分は他の機械と大して違いはない。これらのシステムの多くは、人間から事前に状況や場面ごとの対応を細かく指示され、その行動表のとおりしか行動や学習ができないのだ。しかも新しく知識を入れようとする、既存の知識との間に矛盾を生じることがよくある。それを防ぐため、ここでも人間が細かなプログラミング処理を施さなければならない場合が多い。

ところが人間の乳幼児の場合はどうだろうか。我々は「学習するための詳細な方法」を教わったのだろうか？ 乳幼児は事前知識なしに、周囲の環境や人を観察して知識や概念を構築していく。人にとっては当たり前のことであるが機械にそれを再現させるにはとても困難なことであり、機械が

人に近づくのを妨げる大きな壁である。

科学技術の発展によって、人と機械の関係はとて親密なものになってきている。これからの未来、機械が人と共存していくためには人のような柔軟性、汎用性が必要となってくる。長谷川研究室では画像処理の立場からこのような汎用性をもつ機械を支えるシステムの開発を試みた。そして、色や形の概念を事前知識なしに学習していく画像認識システム、SOINNを作り出した。

SOINNとは「Self Organizing Incremental Neural Networks」の略称であり、ノイズが多く混じった視覚情報にも対応でき、知識の新規獲得に際してエラーを起こさずに継続的かつ追加的に色や形の概念を学習することができる、という特徴を持っている。ノイズが多く混じった視覚情報というのは現実世界の視覚情報を想定しているのであるが、ノイズが多いとはいったいどのようなことなのだろうか。

ここで自分が日当たりの良い桜並木を歩こうとしている場面を思い浮かべてほしい。この場合においてノイズとは、散っている桜の花びらや生い茂る草木、太陽の光などが該当し、それらの余計

な情報が多く道にかぶっていても、人はそれを無視して重要な部分だけ見ることができる。つまり人は情報量の多い視覚情報から重要な部分を抽出し、そこから物の判断や学習をしている。この情報の取捨選択が、プログラムで動いている機械には困難なのだ。長谷川研究室では、この問題点を解決することを念頭に置いて SOINN の開発を進めた。

図1は SOINN と従来のシステムに対して、入力として多くのノイズを含む画像を与え、画像から読み取れる形状を正しくつかめるかの実験結果を比較したものである。SOINN と比較するために従来のシステムの代表として、GNG と GNG-U というものを用いた。まず入力された画像を見てもらいたい。そこには黒点が集まった五つのパーツがあると判断できるだろう。そこで SOINN の出力した結果を見ると、同じく五つのパーツを認識できていることがわかる。しかし従来のシステムの結果では何の形にもなっていないかったり、一部しか学習できていなかったりしている。

従来のシステムではなぜ失敗するのだろうか。この GNG、GNG-U というシステムは、画像から刺激(図1の入力画像の黒い部分)を受けると、Node と呼ばれる点をその場所に置いていく。そ

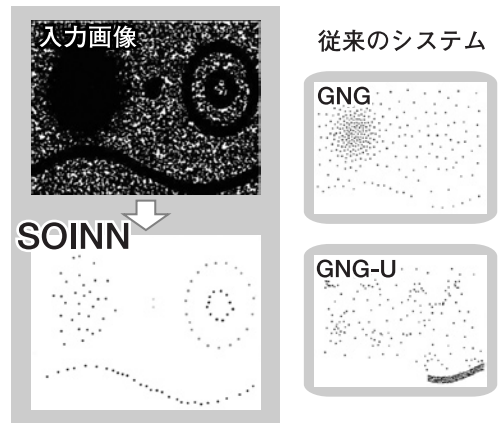


図1 SOINN と従来のシステムの比較

うして刺激のある部分とない部分にわけ、画像の形状的特徴を示すことができる。しかしこのシステムは万能ではなく、実験に使ったようなノイズの強い画像では、特徴をつかむのに unnecessary 部分からも刺激を受けて Node を増やしてしまう。その結果、特徴をつかみづらくなっている。

また従来のシステムには、学習範囲以外の刺激に対してどう反応するかという問題もあった。入力された画像全体としてみると黒点はあちらこちらに見受けられる。そのため学習範囲以外から

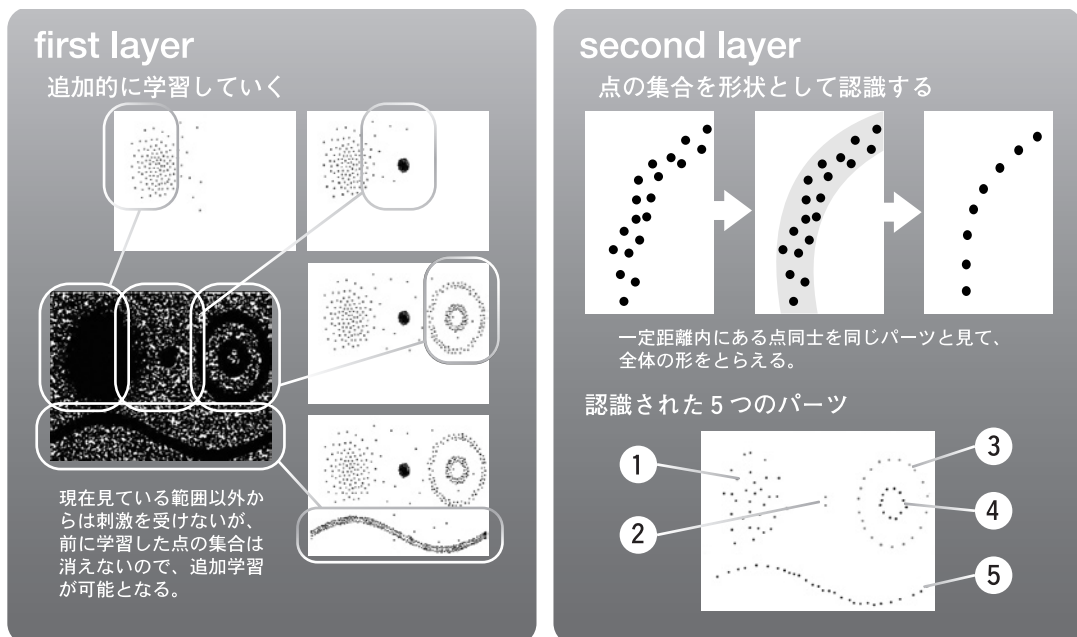


図2 SOINN の画像認識の課程

の刺激も常に受けるようにすると全体に Node が広がってしまい、形を成すことができない。これが GNG の出力結果である。逆に GNG-U では学習範囲以外に黒点があるとしても刺激がないものとして認識させてみたところ、学習範囲以外の Node 数を減らしてしまった。そのため最近学習した部分の情報のみがパーツとして認識されるので、持続的かつ追加的な学習をすることができなかったのだ。

SOINN を作るにあたりこうした従来のシステムの欠点を改善するために、長谷川先生は二つの項目に注目した。それは刺激の密度と自律的な Node 数の管理である。

入力された画像の形状の特徴をつかむ上では黒点の密度が一つの目安となってくる。黒点の密度が大きい領域を一つのパーツとして認識するのだ。SOINN は黒点から刺激を受けて Node を置いていくため、その密度の大きさに比例して置いていく Node の数も増えていく。しかし入力されてくる黒点全てに対して刺激を認識していると、Node の数の差異こそあれど入力された画像全てを一つのパーツとして認識してしまうことになってしまう。そこで黒点の密度、即ち刺激の密度がある一定の値より低い部分を取り除くことで、ノイズの効果的な除去を行うことに成功した。

追加的な学習を可能にするためには、その時々学習している範囲以外の刺激をどう扱うかが重要になってくる。GNG や GNG-U のように学習範囲以外の刺激に対してまで刺激の有無を判断してしまえばパーツ認識が難しくなる。そこで、Node がある一定数まで増えたら情報が集まったと判断し生成を止め、既に集まった Node は残しておくという仕組みを導入した。これによりシステム自身が Node の数の上限や存続を調節する事ができ、持続的な学習を可能とし、追加的に形概念を習得することを可能としたのである。

図2は SOINN が持続的に学習している様子である。この図のようにある一定範囲ごとに Node を置いていき、一つ一つの部分的形状の特徴を学習していく。SOINN には first layer と second layer という二つの層があり、全ての形状の特徴の学習が終了し Node の分布が確定するまでを first layer で行う。そしてその first layer の情報から形概念を抽出し、五つのパーツがあることを、second layer で認識する。

この SOINN は今現在も改良が進められており、その発展形の E-SOINN では first layer を置かずに形概念を直接習得することが可能である。長谷川研究室ではより汎用性の高いシステムを構築している。

## 記録から記憶へ

これまでに紹介した SOINN というシステムは、情報量の多い視覚情報から重要な部分だけを抽出し、色や形概念を習得することができる。長谷川研究室ではこの SOINN と音声の動的計画法というものをを用いてヒューマノイドロボットを実際に作成している。このロボットは、人が物を指しその名前や色や形状を言うと、その情報を覚え、指差してもらったものの特徴を学習することができる。また、覚えたものに関しては、人が指差すとその名前などを音声で応えてくれる。



図3 概念により認識する「レモン」

例えば黄色くて楕円形のを指差して「黄色、楕円、レモン」というと、ロボットはその視覚情報と音声情報の組み合わせから、その個体、「レモン」を認識する。視覚情報からは SOINN でその色や形概念を習得し、それと「レモン」を結びつける。音声情報からは動的計画法という手法を用いて音声波形の特徴的な部分をベクトルデータにし、それを形や色の概念、「レモン」という物に関連付ける。そしてこれらを結びつけることによりロボットの頭の中に「レモン」のイメージが創造されるのだ。これ以降ロボットは、人にそれを指し示されると、その個体に関連付けられた音声のベクトルデータを抽出し、「レモン」と発声する。

一見するとただ物を覚えているだけのようにみえるかもしれないが、学習の根本において、この

ロボットは他の機械とは一線を画している。

例えば電子辞書で「レモン」と入力してみる。すると、電子辞書はレモンに関する詳細な情報をディスプレイに表示してくれる。しかし電子辞書自身はレモンについて理解しているわけではない。レモンと入力されたら決まった0と1の記号列を返すだけで、学習するときはそのレモンに関する情報の記号列を人間に直接与えてもらわなければならない。

しかしこのロボットの場合、色と形状と音声だけではあるが、「レモン」と言われれば、黄色で楕円でレモンであるということを思い描くことができるのだ。しかもこの「レモン」に関する情報は、ロボットがその内容を理解している証拠に、レモンと言われたときに限らず、他の黄色の物や楕円の物のときにも、それが黄色である、楕円であるとわかるのである。例えば茶色で楕円である「ラグビーボール」を学習するとき、このロボットは「レモン」と形が同じことに気付くが、電子辞書にとっては単なる別の記号列である。このロボットは従来の機械が行う「記録」とは根本的に異なる、人と同じプロセス「記憶」を行っているのだ。

しかもこのロボットはたくさんの事例を見せて教えていっても、矛盾が生じないようにプログラミング処理をするなどという面倒なことを必要とせず、持続的に学習していくことができる。むしろたくさんの事例を見聞きさせればさせるほど、色や形概念把握がより明確になる。例えば「三角、楕円、長方形」の三つしか形概念を教わっ

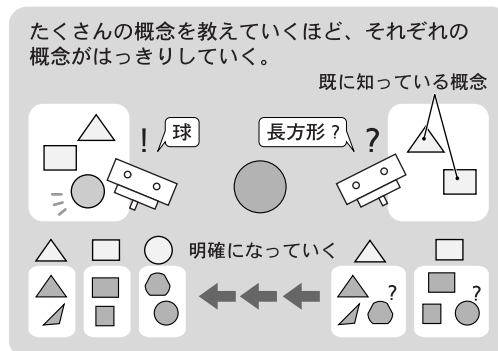
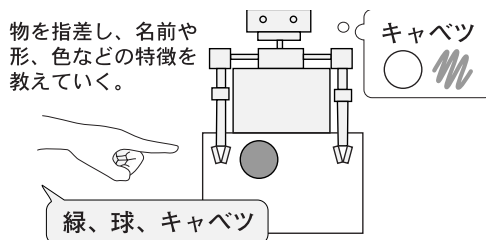


図4 概念の学習

ていなかった場合、ロボットはどんな形でもこの三つのどれかだと判断しようとする。それだけ一つ一つの形概念が曖昧なのだ。ここで「球」という形を学習し、その概念を習得したとすると、今まで楕円や三角と判断していた球らしいものを「球」と識別できるようになり、それだけ「三角」や「楕円」の概念がはっきりしていく。このようにしてこのロボットは継続的、追加的に色や形の概念とそれらの関連付けによる物の特徴の学習をしていくことができる。



## 新たな機能の探求

長谷川先生は他にもこのロボットに様々な機能拡張をしており、動きとそれに関連した文法の学習と既知度とそれを用いた知識欲の表現をロボットに可能にさせた。

まず前者の「動きとそれに関連した文法の学習」について説明する。簡潔にいうと、物体の動きの軌跡をとらえて、その動きを指す動詞と、主語や目的語となる名詞の語順などの文法を学習することである。例を示してみよう。黄色で楕円形の物を「レモン」、緑で球形の物を「キャベツ」と学習させた後、レモンをキャベツに近づけながら「レモン、キャベツ、近づける」という声を聞かせる。

するとこのロボットは「近づける」という動詞の意味を物体同士の位置関係、物体の動きからとらえると同時に、日本語の文法を学習する。ここでわざわざ「日本語の」と言ったのは、英語などの他の言語の文法も学習できるからである。「lemon move near to cabbage」と英語で教えれば、主語→動詞→目的語という英語の語順、文法を学習できるのだ。品詞の並びを判断するボトムアップと、実際にその文法通りに実行したら現実世界でどうなるかを判断するトップダウンという二つの機能で未知の文法を自分で学習している。そしてこのロボットは動作についても、学習した内容を実行

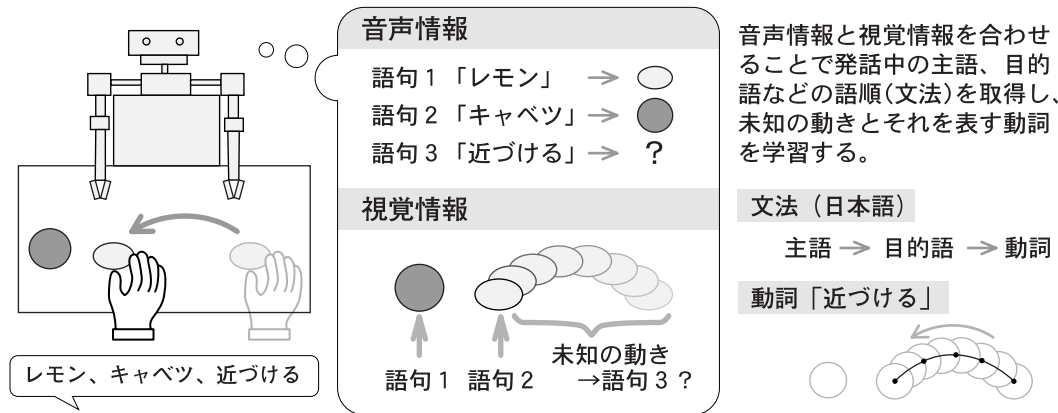


図5 文法と動詞の学習

することができる。「レモン、キャベツ、近づける」と人が指示すると、ロボットは自分のアームを使って実際にレモンをキャベツに近づけてみせる。

次に後者の「既知度とそれを用いた知識欲の表現」について説明する。これは人の場合の好奇心に基づく行動と考えればわかりやすい。人は未知の物体や現象に遭遇したとき、それを知りたいという欲求にかられる。そして周囲の人に聞いたり辞書などで調べたりしてその欲求を満たし多くの知識を習得していく。物心ついた子供などを考えてみればそのことはわかるだろう。このロボットのシステムも同じような機能を持っている。ここでは、与えられた情報に対しそれが既知っている情報かどうかを判断するのに既知度とよばれる統計データを用いている。既知度が高いということはロボットの持つ情報量が多くその情報が確かなものであることを示し、逆に既知度が低いということは情報量が少なくその情報が不確かであることを示す。そしてこのロボットは既知度の低い物体に関する情報の提示を優先的に人間に要求する機能を持っている。これにより色々並んでいる物の中で、既知度の増加分の総和が最大になるような、つまりは学習効果が大きいものを優先的に選ぶことができるのである。人の指示がなければ動こうとしないこれまでの機械とは異なり、この

ロボットは人のように効率的かつ能動的に学習することができるのだ。また、知識欲を持たせて学習させた場合と、特に法則性なくランダムに学習させた場合とで比較して実験してみたところ、知識欲を持たせたほうが学習効率が良いという結果が得られた。このロボットは日々、人に近づいているのである。

長谷川研究室ではこのヒューマノイドロボットに続いてさらに「常識」を学習・獲得できるようなメカニズムの構築を目指している。人が脊髄反射のレベルで認識や判断していることも機械にはまだ難しいのだ。例えば手のひらに卵をのせその手をひっくり返すと、卵が地面に落ちて割れる。そのような簡単な因果関係や、いつも通る通学路の道順などの習慣をその過程の中で常識として発見し、習得するようなロボットを目指して研究している。また「フレーム問題」というものがあり、通常コンピューターは意識すべきものも意識しないで無視していいものも、両方考慮に入れて判断を行う。人は当然判断に必要な要素だけに注目しているのに対して、コンピューターは不必要な要素も含めて判断してしまう。このような情報の取捨の判断も人工知能における重要な課題である。

「もっと人らしいロボットを。」 困難な道のりではあるが、長谷川先生はその路を今確かに一歩一歩進み続けている。

「人らしいロボット」これは筆者自身の夢でもあり今回の取材は何もかもがとても興味深かった

です。お忙しい中取材に大変親切に応じて下さった長谷川先生に御礼申し上げます。（新沼 匠）