

画像処理技術の展望と コグニティブ・コンピューティング Deep Learningへの期待

デジタル画像処理技術には50年以上の研究開発の歴史があり、1966年にはIBMが光学文字読み取り装置(OCR)を開発しました。1970年代には、コンピューター断層撮影法(CT)が登場して特に医学分野に貢献し、1980年代の日本では画像処理用サブルーチンのパッケージが配布されてソフトウェアが充実したことで発展期を支えました。また、コンピューター・グラフィックスを含めたデジタル画像・映像を作り出す技術も大いに発展しました。

画像処理技術に対する人々の大きな期待の一つは、“ロボットの目”を実現することです。本稿では、この10年ほどでおきた画像処理技術の発展と今後の展望について、いくつかの例を取り上げながら解説します。また、IBMの最近の取り組みについてもご紹介します。

▶▶ 1. Introduction

人工知能(Artificial Intelligence)研究の誕生以来、人間が思考する過程をコンピューターで再現して、例えば、ヘルパー・ロボットや思考するマシンが現実のものとなるであろうと多くの人が期待していました。しかし、2014年の今日でも、それを目にするのがないのはなぜでしょうか。

最新のデバイス(マシン)には、さまざまなタスクを実行するのに十分なハードウェア機能があります。しかし、それを支えるソフトウェア機能が不足しています。特に、コントロールと知覚の機能です。ここで、コントロールとは、与えられた制約をすべて満たしながら実行に移す判断を指し、知覚とは、そのような判断や決定をする前段階として周囲の状況をつかむことを指します。

現実の世界は非常に複雑なので、数学的な表現ですべての現象を記述することは不可能です。将来、マシンが人間のように振る舞えるようになるには、自動で知識を取り入れて、ルールや手順などを学習して、また対話をしながらある程度の理解をする必要があるでしょう。膨大な医療データや海洋環境汚染のマルチセンサー・データなど、まだよく解明されていない不透明なデータ・

ソースの中からでも、深い洞察を導き出すことが重要です。そのようなソフトウェアは一部で、いわゆるif-then-elseの形式でプログラム化されるでしょう。しかし、すべての手続きやルールを書き下すことは不要です。代わりに、データからコンピューターが自ら学ぶようになり、知識を増やしたり、変化に追従したりするようになるはずで

前述のコントロールに関しては、機械学習が広く採用されるに至り、人間によってプログラムされたシステムではなく、学習した判断基準に基づいたシステムにより、複雑な道をナビゲートしたり無人のヘリコプターを飛ばしたりできるようになっており、人間ができる範囲を上回ることもあります。

一方で、知覚は、まだまだ未解決の領域です。ロボットに面倒な仕事を任せるための、最後の障害と言えるかもしれません。われわれにはさまざまな知覚の機能がありますが、視覚は其中最も複雑な機能です。人間の脳では、大脳皮質の30%が視覚に使われます。ちなみに、触覚は8%、聴覚は3%です[1]。そのような知覚が、異種のマルチセンサーとして同時に複合的に処理されるのです。



図1. 汎用の特徴表現と道路標識認識への応用

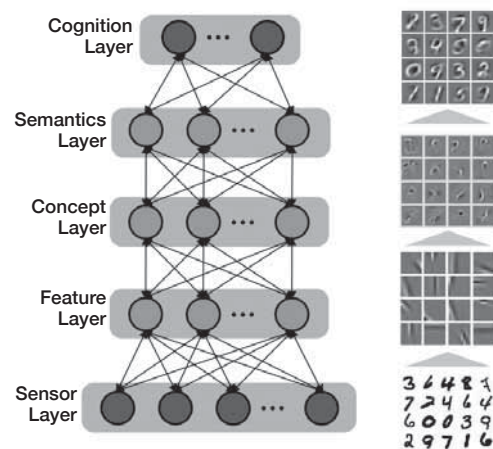


図2. 視覚野の構造に似た積層(左)と情報の抽象化(右:下から上へ)

2. Objective

ロボットの目の機能を実現するためのコンピューター・ビジョンでは、どのような機械学習の手法と組み合わせる場合でも、データ表現(特徴)を適正に選択して設計しなければ期待通りの識別性能が得られません。そのような特徴表現として最も成功した例には、SIFT(Scale-invariant feature transform)やHOG(Histogram of oriented gradients)があります。これらは画像処理の貢献や適用を飛躍的に高めた技術で、例えば、図1に示すように、特徴表現(点)の対応を調べるだけで、シーンの中から目的の標識を見つけることができます。

これらの技術は、画像処理の歴史における一つのターニング・ポイントだったと言えますが、特徴表現として万能というわけではありませんでした。識別能力を高めるための特徴さえも、人間の介在なしに、低レベル(例えば数値レベル)のセンサー情報に隠された説明要素を読み解いて取り入れるような学習機能が実現されれば、人工知能の完成にさらに一歩近づくことでしょう。

さらに次のレベルの期待としては、複数のセンサー情報を学習し融合して利用することで、レーダーのような本来人間の持ち得ないセンサー機能を複合して、人間を超える知覚を持つ人工知能を作ることにも可能となるでしょう。

3. Deep Learning

ニューロン(神経細胞)を模した人工ニューラルネットワークは、層(レイヤー)の積み重ねの構造で学習の

機能を実現しています。霊長類脳は、視覚野において同様の積み重ね構造を持つため、図2の左側に示すような多層のニューラルネットワークはさらに高度な学習モデルを実現できると期待されていました。

ところが、層の数が増えると、人工ニューラルネットワークの層間接続を自動構成するためのトレーニング(学習)が期待通りには完了しません。それが長年にわたり研究者の悩みでした。すなわち、複数の深い層をまとめてトレーニング処理すると、与えた例題に対しては正しい回答(分類結果など)を出力しますが、未知のデータに対しては期待する結果を示さない、いわゆる過学習の状態となります。

しかし2006年頃、Geoff Hinton教授やその他の研究者により、深い層による学習にブレークスルーがありました[2]。ここでのポイントは、前処理としてまず各層に分けて入力した(センサー)信号をうまく弁別するように、例えば、エネルギーを最小化するような数学モデルを導入してトレーニングを行います。各層でのトレーニングを終えたら最後に、全体のニューラルネットワークに対して、入力と期待する出力(答え)の両方を与えて微調整します。各層のトレーニングの段階では、例えば、LとRの音は異なるので弁別しますが、学習の対象課題が日本語の解釈なら、その差は出力にとって重要ではないため、最終的なトレーニングによって簡略化されて消滅するかもしれません。大人になってから他国語を学ぶ際に苦労するのは、案外、このような現象で理由付けできるかもしれません。

さて、Deep Learningは、特にその前段(図2の右下側)において特徴表現を自動的に学習して、弁別機能に取り込むことが期待されます。画像や音声など、入力が生センサーデータに近い分野で、それまでのコンピューター・プログラムによる認識性能の記録を大きく塗り替えて、特に良い結果を出しているのは、この性質のおかげだと考えられています。

▶▶ 4. Research at IBM

IBMリサーチは、コグニティブ・コンピューティングの分野でも技術開発をリードしています。Watson(Deep Q&A)システム[3]は、クイズ番組で人間のチャンピオンを負かす快挙を成し遂げました。IBMはそのようなテキスト文書だけでなく、聴覚や視覚データなどもシステムに取り込んで、音声認識や物体認識などの領域で拡張を続けています。

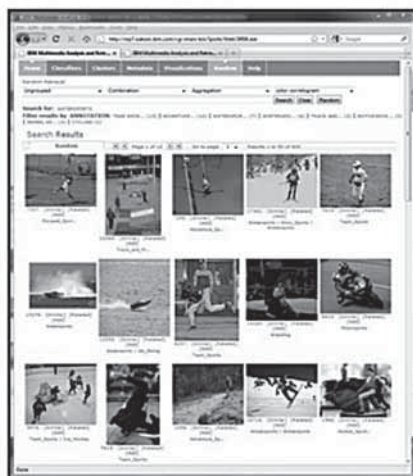
IBMはコグニティブ・コンピューティングの視覚的領域でもいくつか製品およびソリューションを提供しています。

IMARS[4]は、視覚的特徴を生かした機械学習フレームワークであり、画像や映像に意味付けをしてモデル化を行い分類する機能を提供します。多数の分類器を組み込みで用意しているため、図3の右側に示すように、場所、人々、物体、状況、活動、事象などもカバーします。新たに

分類項目を付け加えることも簡単です。

IBM IVA(Intelligent Video Analytics)[5]は、観測した視覚的データから、行動に関する事象や属性、パターンなどを特定します(図4)。このビデオ解析ソフトウェアは、ビデオの連続ストリームを観察しながらリアルタイムでセキュリティー警告を出したり、あるいは撮りためたデータをまとめて処理して、特定の出来事やパターンを拾い出したり、実時間とバッチの両方のモードを提供します。都市の規模で組織的に、あるいはそれをまたがる広い範囲で、データを解析して得られた深い洞察を共有して、よりスマートな決定を導くのに大いに役立つと考えています。状況の流れをとらえた検索インデックスを作成することで、ますます高度化していくセキュリティーの要件にも対応します。

IBM System G[6]は、グラフ構造に関するさまざまな計算ツールを包括するシステムです。“G”はgraphの最初の文字であり、グラフの大小を問わず、固定グラフにも変化するグラフにも対応し、固定属性流でもベイズ統計流でも計算させることができる、とても柔軟なツールキットです。ソーシャル・ネットワークの解析や異常検知、インパクト予測、個人趣向や集団の解析など多岐にわたって利用できるフレームワークです(図5)。視覚的な機能や使い方を挙げるなら、画像データを入力とし



Automatic recognition of sports photos(156 sports categories)

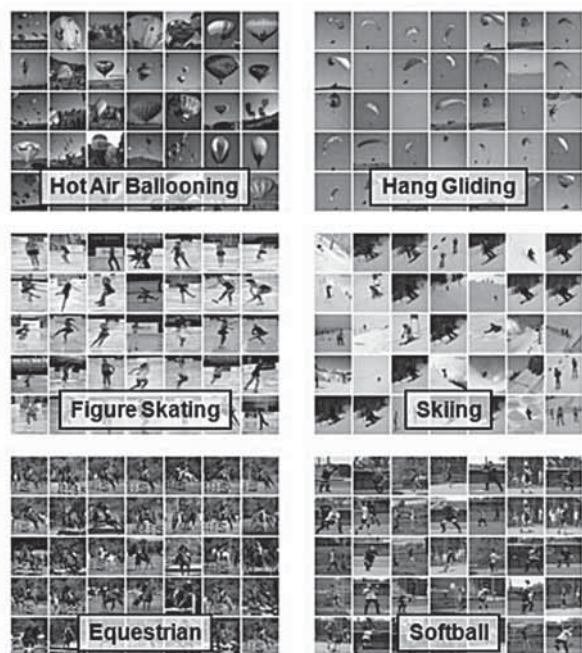


図3. IMARSの分類例

た感情解析にも適用できます。

IBMリサーチではほかにも、販売店舗での不正行為の検出や鉄道の安全確認、テールライト追従といった運転者補助など、幅広いアプリケーションに対応してきました。

▶▶ 5. Future trends and challenges

この10年でコンピューター・ビジョン研究は大きな進展を見せました。しかし、いまだ子供レベルの視覚機能・性能にさえ及ばず、追いつくにはまだまだ長い道のりが必要でしょう。そのために、どのような技術革新が必要なのか厳密には分かりませんが、大体の方向性はすでに明らかになったのかもしれませんが。最先端のDeep Learning手法は人間の脳をまねており、同様の仕組みを構築することは、人間が持つ知覚機能の秘密を暗に利用しているのでしょう。

乗り物の運転者を補助するシステムは、高度な視覚的認知技術を最初に適用する重要な製品となるかもしれません。この変革は、ぶつからないクルマ、事故のないクルマへとつながり、搭乗者を守るために必要なボディの頑強さが不要となり、軽いクルマ、そして低燃費・無公害のクルマへと正方向のスパイラルに導くことも期待できます。

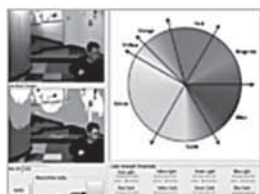
マシンが視覚的認知機能を充実していくことは、



Rich people search capabilities using a person's features and tricolor combinations



Alerts created when abnormal movement of objects outside normal patterns is detected in live stream and historical data



Advanced search features with manual and auto size correction to increase search results



Statistical analysis of specific activity within a timeframe to perform a trend analysis

図4. IBM IVA(人の検索や異常行動の検出など)

人間の作業を代替する上で欠かすことのできない重要で本質的な革新であると言えるでしょう。

[参考文献]

- [1] Lauralee Sherwood, Hillar Klandorf and Paul Yancey; Animal Physiology : From Genes to Organisms, Cengage Learning, 2013
- [2] Hinton, G.E.; Osindero, S.; Teh, Y. "A Fast Learning Algorithm for Deep Belief Nets", Neural Computation, 18-7, 2006
- [3] Watson – <http://www.ibm.com/smarterplanet/us/en/ibmwatson/>
- [4] IMARS – http://researcher.watson.ibm.com/researcher/view_group.php?id=877
- [5] IVA – <http://www-03.ibm.com/software/products/en/intelligent-video-analytics>
- [6] System G – <http://systemg.research.ibm.com/>



日本アイ・ビー・エム株式会社
東京基礎研究所 コグニティブ・コンピューティング
シニア・マネージャー リサーチ

清水 周一
Shuichi Shimizu

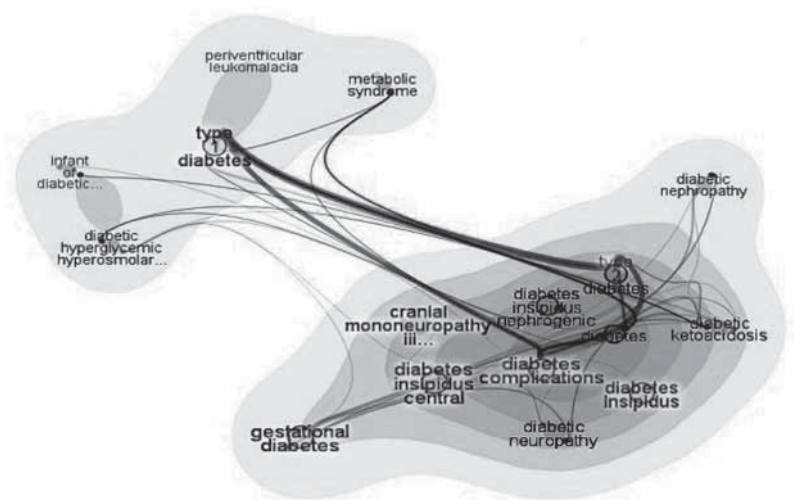
京都大学工学部、京都大学大学院工学研究科修士課程を経て、1987年より日本IBM東京基礎研究所に勤務。現在、コグニティブ・アプリケーション部門のマネージャ。人工知能学会1994年度論文賞受賞。



日本アイ・ビー・エム株式会社
東京基礎研究所 コグニティブ・コンピューティング
主任研究員

ムナワー・アシム
Asim Munawar

パキスタン国立技術情報大学を2003年に卒業し、母国のソフトウェア会社を経て、北海道大学情報科学研究科より修士号および博士号を取得。2012年に日本IBM東京基礎研究所に入所して、現在コンピューター・ビジョンおよび機械学習の研究に従事。



FacetAlas visualizes heterogeneous types of patient relations

図5. IBM System G の適用例