機械学習とディープラーニングの相違点

ジョン・ヴィッカーズ

教師あり学習手法には、畳み込みニューラルネットワークに勝る本質的なメリットがある。

ディープラーニングは最近、マシンビジョンの世界で高い関心を集め、大いに議論されているトピックである。

難しい分類要件が課せられるマシンビジョンに、特によく適している。ディープラーニングは技術用語ではないが、一般的にはニューラルネットワークを使用する。畳み込みニューラルネットワーク（CNN：Convolutional Neural Network）が、商用提供されている画像処理パッケージ内のツールとして提供されるようになったことで、はるかに広範囲にわたるユーザーがこの技術を利用できるようになった。

とはいえ、実際に応用するには、特にトレーニングフェーズにおいて、ビジョン専門家によるかなりの介入がやはり必要である。ニューラルネットワークは、脳のハードウエアまたはソフトウエアモデルであり、単純な意思決定または論理ユニット（ニューロン、パーセプトロン）を、その入力、出力、判断において組み合わせることにより、脳の機能を模倣できる、大規模で複雑な意思決定システム（ネットワーク）を構成する。

単純なレベルでは、トレーニングされていないニューラルネットワークは何も「知らず」、ランダムで無秩序な結果を出力するが、ニューロンとネットワークを「トレーニング」すると、望ましい結果を出力するようになる。システムは自己学習を行うが、その動作は、トレーニングデータセットのサイズおよび品質と、ニューラルネットワークのアーキテクチュアに大いに依存する。他の機械学習手法としては、教師あり学習のカテゴリーに含まれるものが、これに対する有効な代替手段となり得る。

教師なし学習

ニューラルネットワークは、事前知識がなく、すべての知識がトレーニングセットから推論される、教師なし学習に非常に適している。

一般的に、データセットと可能性（解答として想定される可能性の範囲）は非常に大きいので、統計的には解答が存在する。ツールは、データ内の分類（クラス分け）を見つけようとするが、そこに至るまでにかなりの量の学習が必要になり、解答を見出すまでに数多くの「行き止まり」に突き当たる可能性がある

これは、人間が実世界の画像を処理する過程に似ている。人間は、「経験」に基づいて、画像の一部を識別することを学習する。つまり、類似のデータセットを大量に確認し、判断が正しかったかどうかというフィードバックを得ることによって、学習していく。人工的なニューラルネットワークでも、同じことが行われる。大規模なトレーニングセットが必要になるので（一般的に、1クラスにつき約1000枚のトレーニング画像）、許容できるだけの十分な速さでトレーニングを行うために、通常はGPUが必要になるが、一旦トレーニングが完了すれば、ニューラルネットワークによる未知の画像の評価は、CPUを使用して実行できる場合が多い。

図1　CVB Mantoを使用して金属表面の欠陥を検出した様子。

代替手段として、事前にトレーニング済みのネットワークを用意し、少量のトレーニングサンプルでアプリケーションを開発する方法がある。その場合でも、ネットワークの事前トレーニングにおいて、一般的なフィーチャを学習するために、数千枚ものトレーニング画像を含む、大規模で多様なデータセットが必要になる。

教師あり学習

ディープラーニングは、機械学習のカテゴリーのサブセットである。機械学習には、教師あり学習に対応できるアルゴリズムのクラスが含まれる。教師あり学習とは、トレーニングセットにある程度の事前知識があることを意味する。課題は、入力を望ましい出力に接続する関数を作成することで、それには、人間によるある程度の入力が必要になる。

フィーチャの選択と組み合わせが、ど

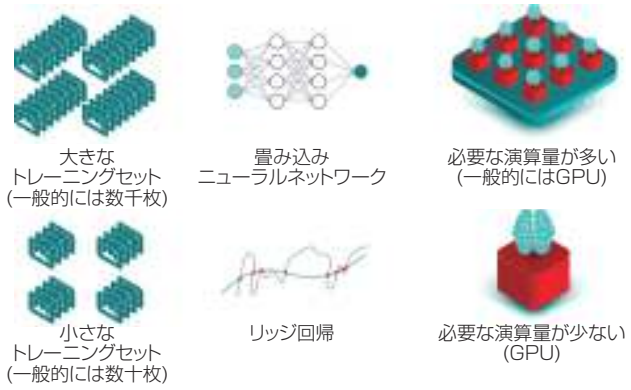

図2　Polimagoは、標準的なパターンマッチング、特にCNNと比べて、比較的少数(一般的に1クラスにつき20～50枚)のトレーニング画像しか必要としない。

れだけ簡単に教師あり学習を達成できるかに影響を与える。従って、最も適切なフィーチャを選択することが不可欠である。ラベル付き入力画像を使用してフィーチャを指定することにより、アルゴリズムが望ましい出力を生み出す関数を生成できるようにする。

サポートベクターマシン、リッジ回帰、決定木、K近傍法、混合ガウスモデルなど、多数の教師あり機械学習手法が利用できる。これらの教師あり学習には、望ましい結果を出力するようにアルゴリズムを方向付けるためのある程度の指示があるため、ニューラルネットワークと比べて、必要なトレーニング画像ははるかに少なくて済む。

教師あり学習手法は一般的に、CNNと比べてトレーニング時に必要となる演算量もはるかに少ない。一般的に短時間でトレーニング可能で、CPUで事足りるので、GPUは不要である。段階的な学習手法が採用される場合が多い。第1段階では、本格的な「学習」が実行される前に、結果の成功レベルが示される。

教師あり学習ツール

上述の教師あり学習手法はすべて、異なるアルゴリズム上のアプローチによって、マシンビジョン用の分類子を確立する。多くの画像処理ライブラリに、これらの手法に基づくツールが含まれている。例えば、「CVB Minos」「CVB Manto」「CVB Polimago」という3つのツールは、英ステマーイメージング社(Stemmer Imaging)が提供するハードウエア非依存のマシンビジョンツールキット「Common Vision Blox」に含まれている。

このようなツールは、CNNで処理できる多数の用途に加えて、マシンビジョンにおける非常に重要なタスクであるパターン認識など、一般的にはCNNでは適切に処理できない一部の用途に対応できる。

決定木

CVB Minosは、決定木をベースとしたパターン認識ツールで、実際のグレースケールフィーチャを使用する。CVB Minosでは、トレーニングフェーズでサンプル画像から自動的に抽出された、グレー値範囲の個別のフィーチャを使用して、オブジェクト認識を行う。トレーニング済みの各オブジェクトタイプの間を区別するフィーチャを検出するだけでなく、さらに確実に正しい検出と分類が行えるように、サンプル画像のネガティブな例も使用する。これにより、2つの類似オブジェクトを区別するとともに、さまざまな背景と混合したパターン(OCRや、紙幣やクレジットカードに用いられるセキュリティプリントなど)にも対応することができる。

サポートベクターマシン

CVB Mantoは、サポートベクターマシン手法を採用し、画像内のフィーチャ(相関、幾何学的なつながり、テクスチャ、色)を使用してオブジェクト

図3　CVB Polimagoでトレイ上のオブジェクトを識別し、その向きと大きさを表示する様子。

図4　車両分類と、シリコンウエハ上の光学文字認識という2つの異なる用途に対し、CVB Polimagoのリッジ回帰アルゴリズムを適用した様子。

表1

カテゴリー	画像情報	トレーニングセットサイズ	分類子数
自動車	240×240px, 8bit RGB	1.7k	6
ウエハOCR	63×68px, 8bit mono	14k	10

表2

性能	RR自動車	RRウエハ	CNN自動車	CNNウエハ
Top-1エラー	5.18	2.02	5.05	5.75
トレーニング時間	0.12min	8min	23min	20min
分類時間	0.048ms	0.045ms	1.722ms	1.288ms

クラスを区別する。組成に有機的なばらつきがあるオブジェクトの識別に適用することができる(異なる種類の果物、人間の顔による性別の判定、表面テクスチャにおける欠陥の特定など)。

Mantoは、トレーニング画像のセットから対象パターンを識別することを学習し、各テスト画像に対する自らの分類判定の確信度を計算する。一般的には、1クラスにつき50～100枚のトレーニング画像が必要で、トレーニングにはCPUが使用される。**図1**は、Mantoを使用して金属表面の欠陥を検出する様子を示したものである。

リッジ回帰

CVB Polimagoは、リッジ回帰という教師あり学習手法を使用する。パターン認識、姿勢推定、オブジェクトトラッキングなどの用途向けに設計されている。一般的なパターンマッチングに対し、正確な分類確率を達成するために、トレーニングフェーズでは、回転、サイズ変更、傾斜などの任意の幾何学変換と、照明の変化を学習する必要がある。

そこで、考えられるすべてのバリエーションを網羅する、トレーニング画像の適切なセットを生成できるかどうかが、課題となる。しかしCVB Polimagoは、特にCNNと比べて、比較的少数(一般的に1クラスにつき20～50枚)のトレーニング画像しか必要としない(**図2**)。

アルゴリズムそのものが、初期トレーニングセットからモデルの合成ビューを生成し、実際のオブジェクトのさまざまな配置や姿勢をシミュレーションする。こうしてアルゴリズムは、数千枚ものトレーニング画像を自動生成することによって、当該パターンのバリエーションを学習し、高い認識率を高速に達成することができる。

リッジ回帰手法の本質的なメリットの1つは、オーバートレーニングの可能性がないことである。最悪の場合で、アルゴリズムは飽和し、それ以上改良できなくなるが、そのような状況になるのは、数千枚ものトレーニング画像を使用した場合のみで、一般的にはそれだけ多くのトレーニング画像が必要になることはない。

ちなみにCNNは、オーバートレーニングの可能性があり、それを回避するための対策が必要である。ロボットで物体をつまむなどのピックアンドプレースや、オブジェクトのトラッキングおよびトレースでは、非常に高い認識率を達成することが重要な基盤となる。**図3**は、Polimagoでトレイ上のオブジェクトを識別し、その向きと大きさを表示する様子を示している。

結果の比較

CNNと教師あり学習ツールは、どちらも以下のような類似の種類の用途に適している。

- 困難な欠陥検出。事前には不明のひっかき傷、くぼみ、キズなど。
- 困難なOCR。手書き、ゆがんだテキスト、スタンプ文字など、大きくばらつきがあるもの。
- 屋外用途。周囲照明によって外観が変わる。
- 有機的な用途。顔、食品、植物など、実際の外観にばらつきがあるもの。

どの用途にどの手法が最も有効かという明確な規則はない。**図4**は、車両分類と、シリコンウエハ上の光学文字認識(OCR)という2つの異なる用途に対する、CVB Polimagoのリッジ回帰アルゴリズムとCNNの性能を、比較して示した図である。

表1に、この比較に使用したトレーニング画像セットと分類子の数に関する情報、**表2**に、トレーニング時間と分類時間の結果を示す。表からは、異なるツールでどれだけの違いが生じ得るかがはっきりと見てとれる。トレーニング後の分類結果の品質も評価して、プロセスが成功しているかどうか、さらなるトレーニングステップや、異なるツールの使用が必要かどうかを判断する必要がある。どのアルゴリズムを使用するかを決めるだけでは、解決策にならない。アルゴリズムの実装が、非常に重要である。

著者紹介
ジョン・ヴィッカーズ博士(Dr. Jon Vickers)は、英ステマーイメージング社(Stemmer Imaging)の技術マネージャー兼製品マネージャー。URL: www.stemmer-imaging.co.uk

VSDJ