

畳み込みの基礎知識

新井 正敏

本章では、畳み込みニューラル・ネットワークのからくりを1からひも解いていきます。1次元のノイズ除去から始め、画像の特徴抽出がどのように行われるかを示します。また、第10章ではフィルタ(カーネル)がニューロンの重みとして扱われ、1つの畳み込みが1ニューロンに相当して学習が行われることまで詳細に解き明かします。

● 畳み込みの意味は「重ね合わせ」と考えよう

そもそも、畳み込みとは何でしょうか。「畳み込む」と「折り畳む」は日本語として似ていますが、文学的な意味合いと技術的な意味合いが異なるため同義ではないようです。本特集では次のように考えて説明をしていきます。

・文学的な「折り畳む」

物の形状やサイズを変えて、よりコンパクトにする行為を指します。例えば、傘を折り畳んで鞆に入れるというように使います。

・技術的な「畳み込む」

数学や情報処理では「畳み込み演算」を指し、特に信号処理や画像処理、ニューラル・ネットワークの分野で「重ね合わせ」を意味します。2つの配列や関数の要素を特定の方法で重ね合わせることで、新しい配列や関数を生成する演算になります。「1つに折り畳む」と言うより「重ね合わせて新たなパターンを得る」になります。

・本特集の「畳み込む」

入力データ(信号や画像など)に対してフィルタ(カーネルとも呼ぶ)を適用する演算です。フィルタの値を入力データに重ね合わせてデータ処理を行うことです。例えば、ノイズ除去、画像エッジ抽出、ぼかし処理、特徴抽出が挙げられます。

図1のように、勉強で本の上にカラー・フィルタ(幅と高さのウィンドウ・サイズ)を置いたことがあるのではないのでしょうか。フィルムを逐次右に移動させていることが畳み込むイメージだととらえると分かりやすいかもしれません。

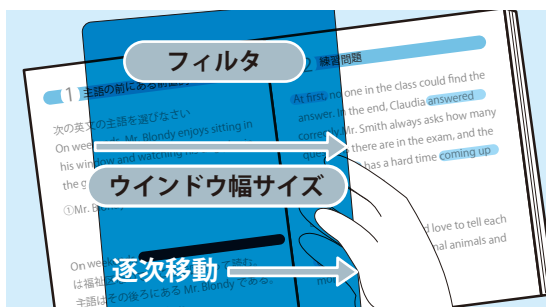


図1 畳み込みのイメージ…勉強で使ったカラー・フィルタを逐次右に移動させるような感じ

● 畳み込みニューラル・ネットワークの全体像

畳み込みニューラル・ネットワーク(CNN; Convolutional Neural Network)は、基本的に畳み込み層(Convolutional Layer)とプーリング層(Pooling Layer)で構成されます。しかし、一般的には図2のように全結合層(fully connected layer)も含まれます。このため、本特集でもCNNは全結合層を含むことになります。CNNの構造は次の通りです。

- ・畳み込み層：入力画像やデータから局所的な特徴を抽出します。畳み込みはフィルタの行列演算によって実現されます。

$$\text{畳み込み出力} = \sum_{i,j} (\text{入力データ} \times \text{フィルタ})_{i,j}$$

- ・プーリング層：特徴マップのデータ数を削減し、計算量を抑えると同時に、重要な特徴を維持します。
プーリング出力 = $\max(\text{フィルタ領域内の値})$
- ・全結合層：畳み込みやプーリングによって抽出された特徴を分類や回帰予測演算を行います。通常はディープ・ニューラル・ネットワークの最後に1つ以上の全結合層があり、最終的な予測を行います。

$$\text{出力} = \text{softmax}(W \times \text{特徴ベクトル} + b)$$

各畳み込み層とプーリング層は活性化関数(主にReLU)を通して出力を行います。この畳み込み層とプーリング層は何重にも接続され、画像の特徴抽出を