深層学習　第 4 回輪講資料

第 6 章　畳込みニューラルネット

2015 年 5 月 28 日

5115F027-7　久保佑介

第 6 章　畳込みニューラルネット

6.1 単純型細胞と複雑型細胞

- **畳込みニューラルネット（convolutional neural network）**
  - 主に画像認識に応用される順伝播型ネットワーク
  - 隣接層間の特定のユニットのみが結合を持つ特別な層を持つ（図 6.1 (b) , (c) ）
    - ✓ これらの層では**畳込み**と**プーリング**という画像処理の基本的な演算を行う
    - ✓ これまではユニットすべてが**全結合 (fully-connected)** されたもの（図 6.1 (a) ）

- 生物の脳の**視覚野（visual cortex）**に関する神経科学の知見
  - **受容野（receptive field）**の局所性
    - ✓ 入力を受け取る感覚器官上の広がり
  - **単細胞細胞（simple cell）**と**複雑型細胞（complex cell）**の存在
    - ✓ 特定の入力パターンのみに反応するが、位置選択性が違う
    - ✓ 単細胞細胞は位置選択性が厳密、複雑型細胞はそれほど厳密でなく、入力パターンが少しずれても反応する

- **物体カテゴリ認識（object category recognition）**
  - 1 枚の画像からそこに写る物体のカテゴリ名を識別する
  - 長い間人には容易だがコンピュータには難しい典型的な問題
  - 多層の畳込みネットは画像認識の問題全般に対する最も重要な技術

6.2 全体の構造

- 入力側から出力側へ向けて、**畳込み層 (convolution layer)** と**プーリング層 (pooling layer)** がペアで順に並び、このペアが複数回繰り返される
  - 畳込み層だけが複数回繰り返された後、プーリング層が 1 層、後に続く場合もある
  - **局所コントラスト正規化層（local contrast normalization, LCN）**が畳込み層とプーリング層の後に挿入されることもある

- 畳込み層とプーリング層の繰り返しの後には、隣接層間のユニットが全結合した層が配置される
  - **全結合層**と呼び、畳込み層などと区別する
  - 全結合層も一般に、複数連続して配置される

- 最後の出力層は順伝播型ネットワークの各層と同じ
  - 目的がクラス分類なら最終の出力層はソフトマックス層となる（p.15）

6.3 畳込み

6.3.1 定義

- 画像とフィルタ
  - 画像
    - ✓ サイズが W×W 画素のグレースケール画像
    - ✓ 画素をインデックス $(i,j)(i=0,\cdots,W-1,j=0,\cdots,W-1)$で表す
    - ✓ 画素(i,j)の画素値をx_{ij}と書き、負の値を含む実数値をとる
  - フィルタ（filter）
    - ✓ サイズが H×H 画素の小さい画像
    - ✓ 画素をインデックス $(p,q)(i=0,\cdots,H-1,j=0,\cdots,H-1)$で表す
    - ✓ 画素(i,j)の画素値をh_{pq}と書き、負の値を含む実数値をとる

- 画像の畳込みは、画像とフィルタ間で定義される積和計算である

$$u_{ij}=\sum_{p=0}^{H-1}\sum_{q-1}^{H-1}x_{i+p},x_{j+q}h_{pq}$$

6.3.2 畳込みの動き

- 画像の畳込み
  - フィルタの濃淡パターンと類似した濃淡パターンが入力画像上のどこにあるか検出（フィルタが表す特徴的な濃淡構造を画像から抽出）する働き

6.3.3 パティング

- 畳込み結果の画像サイズ

$$(W-2\lfloor H/2\rfloor)\times(W-2\lfloor H/2\rfloor)$$

$\lfloor . \rfloor$は小数点以下を切り下げて整数化する演算子とする

- パティング
  - 入力が画像の外側に幅 $\lfloor H/2\rfloor$ のふちをつけることで、出力画像のサイズが入力画像のサイズと同サイズになるようにする

- **ゼロパティング（zero-padding）**
  - ✓　「ふち」の部分の画素値を 0 にセットする
  - ✓　画像の周辺部が自動的に暗くなってしまう

6.3.4 ストライド

- フィルタの適用位置を 1 画素ずつではなく、数画素ずつずらして計算する
  - フィルタの適用位置の間隔を**ストライド（stride）**と呼ぶ

$$u_{ij} = \sum_{p=0}^{H-1} \sum_{q=0}^{H-1} x_{si+p}, x_{sj+q} h_{pq}$$

s：ストライドの値

  - 出力画像のサイズは約$1/s$倍となり、パティングを行う場合、以下のようになる

$$\left(\left\lfloor (W-1)/s \right\rfloor + 1\right) \times \left(\left\lfloor (W-1)/s \right\rfloor + 1\right)$$

フィルタの適用開始位置を$(i, j) = (0,0)$とする場合

- 大きな画像サイズの入力画像を扱うとき、出力側のユニット数が大きくなりすぎるのを防ぐために、2 以上のストライドが使われることがある
  - ストライドを大きくすることは画像特徴を取りこぼす可能性を意味するため、可能なら避けるべき
  - プーリング層でも同様にストライドの考え方が使われる

6.4 畳込み層

- 畳込み層
  - 畳込みの演算を行う単層ネットワーク
  - 実用的な畳込みネットでは、多チャネルの画像に対し、複数個のフィルタを平行して演算を行う
    - ✓　多チャネルの画像とは、各画素が複数の値を持つ画像
  - 縦横の画素が W×W、チャネル数が K の画像のサイズを W×W×K と表す

- 畳込み層が行う計算の詳細（図 6.7）
  - 第 l 層に位置し、直前の第 l-1 層から K チャネルの画像 $z_{ijk}^{(l-1)}(k = 0, \cdots, K-1)$ を受け取る
  - 入力と同じチャネル数 K を持つ M 種類のフィルタ $h_{pqkm}(m = 0, \cdots, M-1)$ を適用する

- 各フィルタについて並列に計算が実行され、それぞれ 1 チャネルの u_{ijm}を出力する
- 各チャネル $\mathrm{k} = (0, \cdots, K-1)$ について平行に畳込みの計算を行った後、
  結果を画素ごとに全チャネルにわたって加算する
  - ✓ 多チャネルのフィルタの場合、チャネルごとの差異を取り出すことができる

$$u_{ijm} = \sum_{k=0}^{K-1} \sum_{p=0}^{H-1} \sum_{q=0}^{H-1} z_{i+p,j+q,k}^{(l-1)} h_{pqkm} + b_{ijm}$$

バイアスはフィルタごとに各ユニット共通 $\left(b_{ijm} = b_m\right)$ とする

- u_{ijm}に活性化関数を適用し、畳込み層の最終的な出力とする

$$z_{ijm} = f\left(u_{ijm}\right)$$

- 入力画像のチャネル数によらず 1 つのフィルタからの出力は常に 1 チャネルになる
  - ✓ 入力の画像が W×W×K のとき、出力は W×W×M となる
  - ✓ ストライドを行う場合は、出力画像のサイズは小さくなる（約$1/s$倍）

- **重み共有（weight sharing, weight tying）**
  - ✓ 出力層のユニット 1 つは（チャネル m の 1 つの画素）は、入力層の
    H×H×K 個のユニットとのみ結合する
  - ✓ 結合の重みがh_{pqkm}となる（m は今考えているチャネル）
  - ✓ この重みは、出力層の同一チャネルの全ユニットで同じである（共有される）
  - ✓ 結合の局所性と重みを共有することが、畳込み層の特徴

- パラメータ最適化
  - 畳込みネットでも順伝播型ネットワークと同様、勾配降下法による最適化を行う
  - 最適化の対象となるパラメータはフィルタ（およびバイアス）となる
  - 誤差逆伝播によって畳込み層のデルタ、フィルタの誤差勾配が計算される（第 6.7 節）

6.5 プーリング層

- W×W×K の入力画像上で、画素 (i,j) を中心とする H×H 正方領域をとる
  - この中に含まれる画素の集合をP_{ij}で表す
  - P_{ij}内の画素について、チャネル k ごとにH^2個ある画素値からu_{ijk}を求める
  - 出力画像のチャネルは入力画像のチャネルに一致する

- プーリングの方法
  - **最大プーリング（max pooling）**
    - ✓ H^2個の画素値の最大値を選ぶ

$$u_{ijk} = \max_{(p,q)\in P_{ij}} z_{pqk}$$

  - **平均プーリング（average pooling）**
    - ✓ H^2個の画素値の平均値を計算する

$$u_{ijk} = \frac{1}{H^2} \sum_{(p,q)\in P_{ij}} z_{pqk}$$

  - **Lp プーリング（Lp pooling）**
    - ✓ P=1 で平均プーリング、P=∞で最大プーリングを表現できる

$$u_{ijk} = \left(\frac{1}{H^2} \sum_{(p,q)\in P_{ij}} z_{pkq}^{P} \right)^{\frac{1}{P}}$$

- プーリング層のユニットにも活性化関数を適用することは理屈上は可能だが、
  普通は適用しない

$$z_{ijk} = u_{ijk}$$

- パラメータ
  - 結合の重みは、畳込み層のフィルタのように調節可能なものでなく、固定されている
  - プーリング層には学習によって変化するパラメータはない
  - 誤差逆伝播法を実行するときは、デルタの逆伝播計算のみ行う（第 6.7 節）

- プーリング層の処理自体は単純であるが、不明なことも残されている
  - プーリング層が多層の畳込みネットの中で果たす役割
  - 最大プーリングや平均プーリングなどの種類と効果の違い

6.6 正規化層

6.6.1 局所コントラスト正規化層

- 画像認識では、入力画像の全体的な明るさやコントラストの違いを吸収する必要がある
  - カメラで撮影されるとき、カメラの露出（シャッタースピード、絞り、イメージセンサのゲインなど)、環境の照明で大きく変化する

- 正規化の方法
  - 画像の集合（訓練データ）についての統計量を揃える処理
    - ✓ 正規化（3.6.1 項)、白色化（5.5 節）がこれにあたる
      1. 学習画像の画素ごとの平均を求める

$$\tilde{x}_{ijk} = \sum_{n=1}^{N} x_{ijk}^{(n)}$$

$x_{ijk}^{(n)}$は n 番目のサンプルの画素(i,j)のチャネル k の値を表す

      2. 画像からこの平均を差し引いたものを畳込みネットの入力とする

$$x_{ijk} \leftarrow x_{ijk} - \tilde{x}_{ijk}$$

  - **局所コントラスト正規化**
    - ✓ 画像 1 枚 1 枚に対し個別に行う処理
    - ✓ **減算正規化**、**除算正規化**の 2 つがある

6.6.2 単一チャネル画像の正規化

- 単一チャネルの画像x_{ij}に対し、画素 (i,j) を中心とする H×H の正方領域 P_{ij} を考える
- 減算正規化
  - 入力画像の各画素から、P_{ij}に含まれる画素の濃淡値の平均 $\bar{x}_{ij} = \sum_{(p,q)\in P_{ij}} x_{i+p,j+q}$ を差し引く

$$z_{ij} = x_{ij} - \bar{x}_{ij}$$

  - 差し引く平均 $\bar{x}_{ij}$ には重み付き平均を使うこともある

$$\bar{x}_{ij} = \sum_{(p,q)\in P_{ij}} w_{pq} x_{i+p.j+q}$$

    - ✓ その場合、w_{pq} は以下の式を満たし、領域の中央で最大値をとり、周辺部へ向けて低下するようなものとする

$$\sum_{(p,q)\in P_{ij}} w_{pq} = \sum_{p=0}^{H-1} \sum_{q=0}^{H-1} w_{pq} = 1$$

- 除算正規化
  - 同じ局所領域内で画素値の分散を揃える操作
  - 領域 P_{ij} 内の画素値の分散

$$\sigma_{ij}^2 = \sum_{(p,q)\in P_{ij}} w_{pq}\left(x_{i+p,j+q} - \bar{x}_{ij}\right)^2$$

  - 減算正規化を施した入力画像を、標準偏差で割り算する
    - ✓ 濃淡変化が少ない（コントラストが小さい）局所領域ほど濃淡が増幅される

$$z_{ij} = \frac{x_{ij} - \bar{x}_{ij}}{\sigma_{ij}}$$

    - ✓ 入力画像のコントラストが大きい部分にのみ適用する（c：定数）

$$z_{ij} = \frac{x_{ij} - \bar{x}_{ij}}{max\left(c, \sigma_{ij}\right)} \left(\sigma_{ij} < c\right)$$

    - ✓ 同様の効果が σ_{ij} に対して連続的に変化する

$$z_{ij} = \frac{x_{ij} - \bar{x}_{ij}}{\sqrt{c + \sigma_{ij}^2}}$$

6.6.3 多チャネル画像の正規化

- 多チャネルの画像では、上の正規化をチャネルごとに適用することもできるが、チャネル間の相互作用を考える
  - 画素値の平均と分散を求める対象が、全チャネルのわたる局所領域 P_{ij} 内の画素の集合にかわる

- 減算正規化
  - K チャネルからなる画像 x_{ijk} を対象とするときの、重み付き平均

$$\bar{x}_{ij} = \frac{1}{K}\sum_{k=0}^{K-1}\sum_{(p,q)\in P_{ij}} w_{pq} x_{i+p,j+q,k}$$

  - 画素 (i,j) ごとに違うが、チャネル間で共通の $\bar{x}_{ij}$ を差し引く

$$z_{ijk} = x_{ijk} - \bar{x}_{ij}$$

- 除算正規化
  - 画像の全チャネルにわたる局所領域の　P_{ij} の分散

$$\sigma_{ij}^2 = \frac{1}{K}\sum_{k=0}^{K-1}\sum_{(p,q)\in P_{ij}} w_{pq}\left(x_{i+p,j+q,k} - \bar{x}_{ij}\right)^2$$

  - 除算正規化は以下のようになる
    - ✓　単一チャネルと同様に分母を $\sqrt{c+\sigma_{ij}^2}$ とすることもできる

$$z_{ijk} = \frac{x_{ijk} - \bar{x}_{ij}}{max\left(c, \sigma_{ij}\right)}$$

6.8 勾配の計算

- 重み行列 $\mathbf{W}$ を適当に定義すると、順伝播型ネットワークの中間層と同様に表せる
  - $\mathbf{W}$ はサイズ H×H×K の M 個のフィルタの係数h_{pqkm}を、畳込みを再現するように規則的に並べたもの（多くの成分が 0 となる疎行列）

$$\mathbf{u}^{(l)} = \mathbf{W}^{(l)}\mathbf{u}^{(l-1)} + \mathbf{b}^{(l)}$$
$$\mathbf{z}^{(l)} = \boldsymbol{f}^{(l)}(\mathbf{u})$$

- 逆伝播の計算も全結合層の場合と基本的には同じ
  - 同じフィルタの係数が何度も$\mathbf{W}^{(l)}$に現れる（重み共有がある）ことを考慮する
  - フィルタの係数h_{pqkm}から$\mathbf{W} = \mathbf{W}^{(l)}$を作る過程を次のように表す
    1. フィルタの係数h_{pqkm}を適当な順に並べ、H×H×K×M のベクトル $\mathbf{h}$ を作る

    2. $\mathbf{h}$ と同じ長さを持ち、$\mathbf{h}$ と内積をとると、$l-1$層のユニット i と l 層のユニットで j 間の重みw_{ji}を与えるベクトル$\mathbf{t}_{ji}$ を定義する
       - ✓　$\mathbf{t}_{ji}$ は高々1 つの成分が 1 をとり、それ以外は 0 となるベクトル

$$w_{ji} = \mathbf{t}_{ji}^{\mathrm{T}}\mathrm{h}$$

       - ✓　$\mathbf{t}_{ji}$ の成分をt_{jir}と書き、r を固定したときt_{jir}を(i,j)成分に持つ行列を$\boldsymbol{T}_r$と書く

    3. この層 l のデルタを$\boldsymbol{\delta}^{(l)}$と書き、全結合層に対する勾配計算の式を適用すると、$\mathbf{W} = \mathbf{W}^{(l)}$の勾配は以下の式で与えられる

$$\partial\mathbf{W} = \boldsymbol{\delta}^{(l)}\mathbf{z}^{(l-1)\mathrm{T}}$$

4. フィルタの係数 $\mathbf{h}$ についての勾配$\partial\mathbf{h}$に変形する必要がある
   - ✓ $\mathbf{W}$の多くの成分はもともと 0 であり、そうでない成分も重み共有により、同じ変数（フィルタの係数）に対応するため

$$(\partial\mathbf{h})_r = \sum_{i,j} (\mathbf{T}_r \odot \partial\mathbf{W})_{ji}$$

⊙は行列の成分ごとの積、和は行列の全成分の和

5. 層 l の デルタは層 l+1 のデルタ$\boldsymbol{\delta}^{(l+1)}$、層 l+1 について同様に定義された重み行列 $\mathbf{W}^{(l+1)}$を用いて次のように計算される

$$\boldsymbol{\delta}^{(l)} = f^{(l)\prime}(\mathbf{u}^{(l)}) \odot (\mathbf{W}^{(l+1)\mathrm{T}} \boldsymbol{\delta}^{(l+1)})$$

- プーリング層には学習の対象となるパラメータはないため、勾配を計算する必要はないが、下の層に伝えるデルタの逆伝播計算は必要になる
  - プーリングの種類ごとに$\mathbf{W}^{(l+1)}$を定め、上の式を計算する
  - 平均プーリングでは、、層 l+1 のユニット j のサイズ H×H の受容野を P_j として以下のようにする

$$w_{ji}^{(l+1)} = \begin{cases} \dfrac{1}{H^2} & if\ i \in P_j \\ 0 & otherwise \end{cases}$$

  - 最大プーリングでは、層 l+1 のユニット j の同受容野内で、順伝播計算時に最大値を与えた入力層のユニット i のみ$w_{ji} = 1$とし、それ以外を 0 とする

6.8 実例：物体カテゴリ認識

- 物体カテゴリ認識をとりあげ、畳込みネットの適用例を紹介する
  - ILSVRC（ImageNet Large Scale Recognition Chaleenge）というコンテスト

- 入力画像 1 枚に写る、舞台のカテゴリ（クラス）を 1000 種類の中から答える問題
  - 訓練データ：サンプルは各カテゴリ約 1000 枚×1000 種類 = 百万
  - ランダムに取り出した 100 個を表 6.1 に示す

- ネットワークの構成（表 6.2, 図 6.11）
  - 2012 年の ILSVRC で優勝した畳込みネットとほぼ同じもの
  - 5 つの畳込み層、3 つのプーリング層、2 つの局所コントラスト正規化層、3 つの全結合層
  - 学習で決定するパラメータがあるのは、畳込み層と全結合層のみ
  - fc6 と fc7 の層のユニットには第 3.5.3 項のドロップアウト（p=0.5）を適用する

- 確率的勾配降下法を実行したときの学習曲線（図 6.12）
  - ミニバッチのサイズを 128 とする
  - 約 200,000 ミニバッチ（2 万×128 / 約 25 エポック）ほどで収束している
  - 訓練サンプルの画像は、強制的に 256×256 に直した後、中央の 227×227 部分を切り出し、平均画像（図 6.13 (b) ）を差し引く（テストも同様）

- 学習後の畳込みネットへの画像の入力
  - 図 6.13 (a) の画像を入力したときの、畳込み・プーリング・正規化各層の出力を入力側から順に図 6.14~図 6.19 に示す
  - その後の全結合層の出力をプロットしたものを図 6.20 に示す
  - 図 6.20 (a) の fc6 層、(c) の fc7 層とも、活性化の様子はそれほど疎ではない
  - 図 6.20 (e) の fc8 層の出力、特に (f) のソフトマックス関数適用後は、入力画像のカテゴリに鋭いピークが立っている

- 訓練データにない新しい画像を入力したテストの結果（図 6.21）
  - すべての画像で出力されたカテゴリ尤度（ソフトマックス出力）上位 5 位中に、正しいカテゴリが入っている
  - 認識性能が高いだけでなく、間違いを含めた振る舞いが人の感覚に近い
    e.g.) 'lion' や'zebra' は、高い確信度で認識できている
      - ✓ 人が見ても簡単で見間違えようのない画像

    e.g.) 'japanese spaniel'を 'papillon' 、'crane' を"snowplow'と間違えている
      - ✓ 後の 2 つは、カテゴリ分け自身が微妙な人にとっても分類が難しい