ISS-SP-222

# hw/sw 複合体による大規模ニューラルネットワーク実装の検討

鈴木 章央† 堀 三晟† 関根 優年†† 田向 権†

†九州工業大学大学院生命体工学研究科 ††東京農工大学大学院工学府

## 1. はじめに

深層学習という学習方法が様々なコンテストで賞を取り注目を浴びている. これはその名の通りニューラルネットワークを何層にも積み重ねて大規模にしたものである.層を増やすことで学習精度を上げることが出来るが, 既存の手法であるフィードバック学習では層が増えることで学習に時間がかかってしまう. そこで, 事前学習を行い元データから特徴を抽出し, 実際の学習には圧縮したデータを使用することで精度と速度の両方を向上させることに成功した. また, 事前学習には大規模データベースを用いることで精度の向上を図った. しかし, 深層学習を行うためには膨大で複雑な計算量が必要である. そして, この処理をすべてソフトウェアのみを用いて行うと非常に時間がかかる.

## 2. Stacked Denoising Auto Encoder

事前学習の一つとして AutoEncoder(以下:AE)を用いる手法がある [1]. AE は入力層, 隠れ層, 出力層の三層からなるニューラルネットワークである (図 1).

入力層から隠れ層へとエンコードし, 隠れ層から出力層へとデコードする. それぞれ次のように表現される.

$$y = s(Wx+b) \quad (1)$$

$$z = s(W'y+b') \quad (2)$$

$$s(a) = \frac{1}{1+e^{-a}} \quad (3)$$

$x$ は入力層ノードの値, $y$ は隠れ層ノードの値, $z$ は出力層ノードの値である. $W$, $W'$, $b$, $b'$はそれぞれノードに対する重み, ノードへのバイアスを表し, $W'$は$W$の転置行列 $W^T$ で表現される. 入力層と出力層の値を近づけるように学習を繰り返し, パラメータを更新していく.

AE の応用の一つとして, Denoising AutoEncoder(以下:dA)があげられる. dA では入力を確率的に書き換える. ノイズを乗せた入力を再現させるように学習を行うことで学習をより頑強にすることを目的としている.

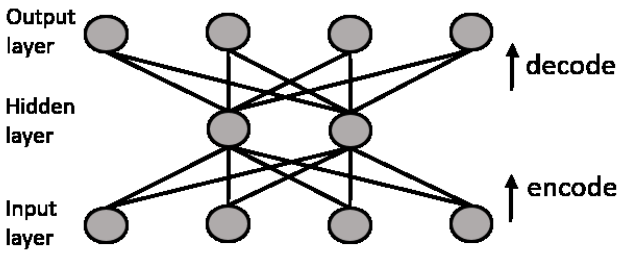


図 1: AutoEncoder

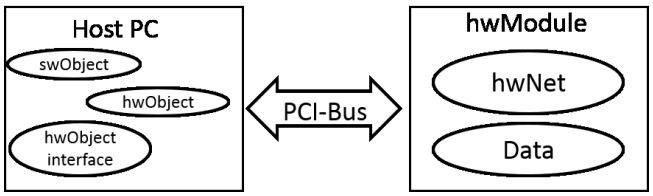


図 2: hw/sw 複合体

また, この AE を何層にも積み重ねたものを Stacked AutoEncoder(以下:SA)という.SA では, 入力層に近い層からパラメータの学習を行っていき, 一層層の学習が終わった後に注目層の隠れ層の値を次の層への入力とし,順次学習を行っていく. SA において通常の AE ではなく dA を積み重ねたものを Stacked Denoising AutoEncoder(以下:SdA)という.

## 3. hw/sw 複合体

深層学習の問題解決のために, hw/sw 複合体[2]を用いることを提案する(図 2). ハードウェアには並列演算が可能であるという利点, ソフトウェアには複雑な演算が可能という利点がそれぞれある. そこで, ハードウェアに加算や減算といった単純な並列演算を, 乗算のような複雑な演算をソフトウェアに行わせることで, 両方の利点を活かし, この問題を解決する. このとき, ハードウェアによる処理をカプセル化し 1 つのモジュールとして設計し, ソフトウェアのライブラリのように容易に使用することができるようにする. また, ハードウェアには低消費電力の FPGA を用いることで, 電力消費を抑えることが出来る.

## 4. まとめ

深層学習は多層ニューラルネットとワークと大規模データベースによって高精度の学習を実現した. しかし, 深層学習を行うためには膨大な量の演算が必要になる. hw/sw 複合体を用いることで高速処理を可能にしていく. ソフトウェアのみで処理を行ったときとの実行時間を比較し, hw/sw 複合体を用いることの有用性を実証していく.

## 参考文献

[1] Pascal Vincent, et al., “Stacked Denoising Autoencoder: Learning Useful Representations in a Deep Network with a Local Denoising Cirterion,” Jounal of Machine Learning Research 11, pp.3371-3408, 2010.

[2] 田向権, 関根優年, “ニューラルネットワークのハードウェア実装とそのシステム化へのアプローチ,” 日本神経回路学会誌, Vol.20, No.4, pp. 166-173, 2013.