

# FPGA に実装した Ternary Sparse XNOR-Net の MNIST 推論

黒木海斗 河原 尊之  
東京理科大学工学部 電気工学科

## 1. はじめに

今日, IoT 化が急激に普及しつつある中で, 組み込み機器に AI をどのように搭載するかが大きな課題となっている[1][2]. ディープラーニングは, 一般的に回路化する際に積和回路が大量にあるほど認識精度が高まるが, その弊害として回路規模や消費電力が増加が発生し, 組み込み機器への搭載が困難になる. その解決策として, 入出力と重みを二値化する XNOR-Net や, 入出力を二値化し重みを三値化する Ternary Sparse XNOR-Net が存在する. そこで, 本研究では Ternary Sparse XNOR-Net を FPGA に実装しシミュレーションを行うことで, その認識精度と回路規模について検証する.

## 2. ニューラルネットワーク

ニューラルネットワークの回路化においては様々な手法が存在するが, ここでは重みの二値化を用いる手法と三値化を用いる手法を紹介する.

### XNOR-Net [3]

重みと入出力をどちらも(-1,1)しかとらない二値とし, 全ての-1を回路上では0と置き換えることで, 乗算をXNOR回路で実現する. 従来最も回路規模を小さくすることができる手法とされている.

### Ternary Sparse XNOR-Net [4]

XNOR-Netと三値化を組み合わせた手法である. 図1にイメージ図を示す. 入力(1,-1)で重みは(-1,0,1)を取るが, 0の重みは配線せず, 重み-1を回路上では0とする. この処理により, 回路規模をXNOR-Netより小さくしつつ, 三値として学習させた重みを用いることで高い認識精度を実現している.

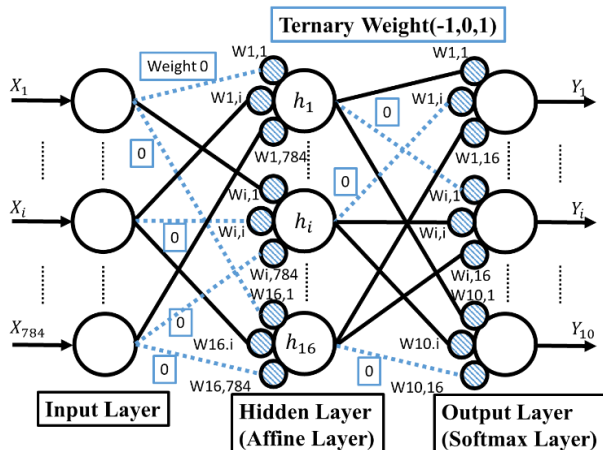


図1. Ternary Sparse XNOR-Net のイメージ図 [4]

## 3. シミュレーション方法

FPGA(Virtex-7,xc7vx485tffg1761-2)への実装とシミュレーションを行う. 対象となる回路は二種類ある. 一つ目は入力が(1,0)で重みが(-1,0,1)のスパース化なしの回路(以下, スパース化なしと記す). 二つ目は図1の Ternary Sparse XNOR-Net である. この二つの回路に MNIST の画像を二値化したものを 60000 枚入力し, その識別精度を測定する.

## 4. シミュレーション結果

表1に二値化した MNIST を 60000 枚入力したときの, スパース化なしと Ternary Sparse XNOR-Net の精度と LUT 使用率を示す.

表1. それぞれの回路における精度と LUT 使用率

回路名	精度[%]	LUT使用率[%]
スパース化なし	88.6	2.36
三値化スパースXNOR-Net	88.4	1.55

表1より, Ternary Sparse XNOR-Net にすることで, 約 0.23%精度が低下するが, LUT 使用率は約 34%減少していることがわかる. このことから, 精度をほぼ低下させることなく, 回路規模を 2/3 程度に抑えられていることが確認できる.

## 5. まとめ

FPGA に Ternary Sparse XNOR-Net を実装することができた. これにより, 精度をほぼ下げることなく回路規模減少を実現できた. 今後は精度低下を無くしつつ, より効果的に回路規模を減少できる手法を検討していきたい.

### 参考文献

- [1] Y. Fukuda et al., "Robustness Evaluation of Restricted Boltzmann Machine against Memory and Logic Error", IEICE Transactions on Electronics, E100.C, Dec 2017, p. 1118-1121.
- [2] R. Oiwa et al., "Timber Health Monitoring using piezoelectric sensor and machine learning", Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA) 2017 IEEE International Conference, Jul 2017, 10.1109/CIVEMSA.2017.7995313.
- [3] M. Rastegari et al., "XNOR-Net: ImageNet Classification Using Binary Convolutional Neural Networks", Aug 2016, arXiv preprint arXiv:1603.05279v4.
- [4] 吉田康彦,河原尊之,"IoT 向け三値化ニューラルネットワークの検討",電子情報通信学会研究会(2017.12)