

再帰型ニューラルネットワークを用いた学習・推論器の FPGA 実装および音声識別 Implementation of learning and inference circuits for recurrent neural network into FPGA and its application to speech recognition

星 郁雄[†] 天野 洋[†] 橋本 大志[†] 田中 大智[†] 下馬場 朋禄[†] 角江 崇[†] 伊藤 智義[†]
Ikuro Hoshi Hiroshi Amano Hiroshi Hashimoto Daichi Tanaka Tomoyoshi Shimobaba
Takashi Kakue Tomoyoshi Ito

1. はじめに

近年、深層学習についての研究が盛んに行われている。その中でも、再帰型ニューラルネットワーク(RNN)はネットワーク内に再帰構造を持つ深層学習の一種で、時系列データの扱いに優れていることから音声認識や自然言語処理の分野で高い成績を残している。

RNNに限らず深層学習では、高い性能の反面計算コストが高いため、クラウドコンピューティングやGPUを用いて計算を処理する研究が進められている。しかし近年では応答性の観点からこの計算をクラウドコンピューティングのように別の場所へ送信し計算するのではなくエッジ側での処理が求められるようになってきている。そこで本稿では、回路を自由に書き換えることができる Field Programmable Gate Array(FPGA)に RNN の推論、学習器を実装し、回路の評価と、実際に音声識別を行った結果の評価までを行う。

2. 実装したネットワーク

3. 推論部

図1に実装した RNN 推論部のネットワーク構造を示す。隠れ層への入力として、入力層のほかに過去の隠れ層出力が入力として存在し再帰構造を形成している。この構造を持つことによりあるデータが入力された際、それ以前に入力されたデータの情報も考慮することができるので時系列データを扱うことが可能になっている。

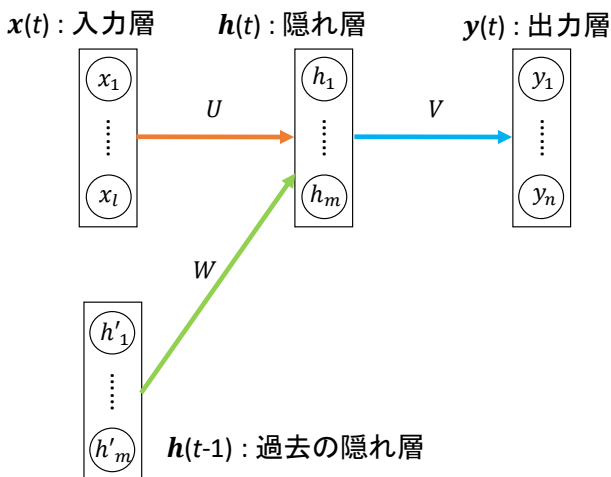


図 1 RNN 推論部

[†] 千葉大学大学院工学研究院 Graduate School of Engineering, Chiba University

3.1 学習部

近年は RNN の学習方法として、推論結果と教師データから求まる誤差を時間ごとにさかのぼり、時間長あたりの誤差を求める Back Propagation Through Time(BPTT)法[1]が一般的とされている。しかしこの手法は FPGA 実装の際、学習に用いる誤差を求めるための数値をすべて時間長分だけ記録しておかなければならず、メモリ使用量が非常に大きくなってしまいう問題点がある。そこで本稿では Real Time Recurrent Learning(RTRL)[2]法を採用した。この学習方法はリアルタイムで学習を行うことに適していて、時間をさかのぼって学習を行う BPTT 法とは異なり、単位時間ごとに推論と学習をそれぞれ行うことで未来へ誤差を伝播させて学習を行う。単位時間ごとに学習することができ時間長分の数値を記録しておく必要がないのでメモリ使用量の削減につながる。図 2 にそれぞれの学習方法を比較した図を示す。

4. 回路設計

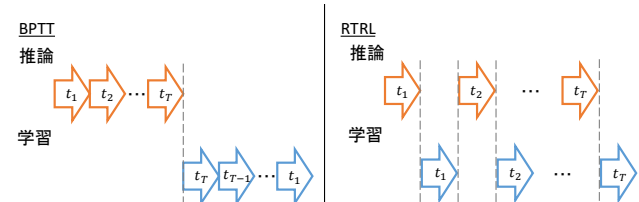


図 2 BPTT 法と RTRL 法の比較

設計した回路の概略図を図 3 に示す。構成としてホスト PC からデータを入力し、データ中に教師データが存在する場合は推論と学習を行い、存在しない場合は推論のみを行い、それぞれの場合での推論結果をホスト PC に返す回路を設計し、FPGA に実装した。

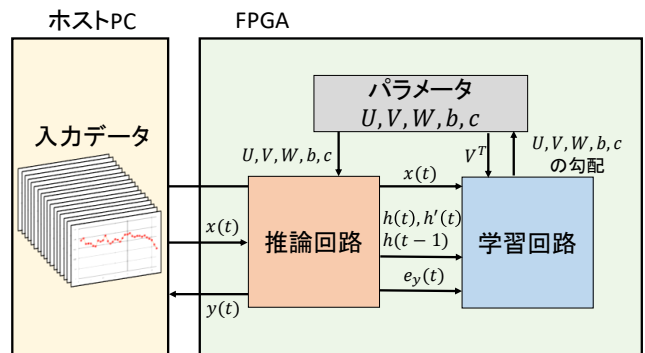


図 3 回路の概略図

また、図 4(a)に設計した回路の推論部ブロック、図 4(b)に設計した回路の学習部ブロックをそれぞれ示す。推論部では入力の値と過去の隠れ層の値から推論を行い、その結果をホスト PC に返す。推論部は、教師データとの誤差、活性化関数の微分と隠れ層の出力をパラメータとして学習部に渡している。活性化関数は、隠れ層に ReLU 関数を、出力層に softmax 関数を用いている。

学習部では入力データと推論部から受け取ったデータを用いてそれぞれの勾配を計算している。本稿では小数点以下 12bit で演算を行っている。単位時間あたりのデータにつき学習を行う場合、誤差が小さくなり勾配が消失してしまうので、十分更新できる大きさになるまで出力直前の積算器で勾配を加算し続けることで更新値を求める。最後に図 4(c)に示した回路を用いてパラメータを更新する。

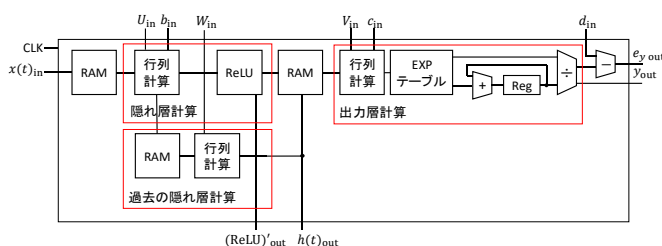


図 4(a) 推論部のブロック図

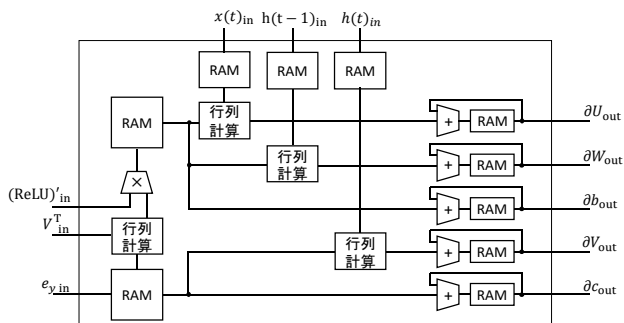


図 4(b) 学習部のブロック図

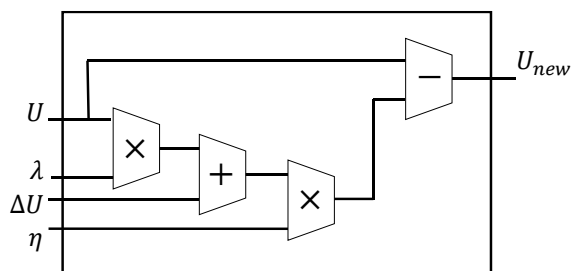


図 4(c) 更新部のブロック図

5. 評価

設計した回路を、回路リソースの使用率と、実際に音声識別を行った場合の結果から評価を行った。使用した FPGA は Xilinx Artix-7 XC7A100TCSG324-2 である。論理合成には Vivado2017.4 を用いた。音声識別には「0」～「9」からなる音声のスペクトルデータを 16 の時系列データに分割したものをを用いた。単位時間当たりのデータ点数は 32 点となっている。中間層ユニット数は 32 とした。

表 1 に回路の仕様リソースを示す。使用可能リソースが少ない Artix-7 でも使用可能リソースの範囲内で実装が可能であることを確認した。

表 1 リソース使用率

リソース	搭載数	使用数	使用率
LUT	63,400	7,143	11.27
LUTRAM	19,000	1,755	9.24
FF	126,800	5,718	4.51
BRAM(36kB)	135	3	2.22
DSP	240	38	15.83

設計した回路の最大動作周波数は 105.8[MHz]となり、動作周波数 100[MHz]で動作させた。

また、図 5 に実際に FPGA を使った音声識別の学習時の正答率変化を示す。グラフから正答率がエポック数を重ねるごとに変化していき、全体で見ると正答率が上昇している様子が確認できた。このグラフから、設計した回路が学習に成功し、またその時の正解を出力できていることがわかる。

最高正答率については、エポック数が 31 回目の場合で 82%の正答率を記録した。

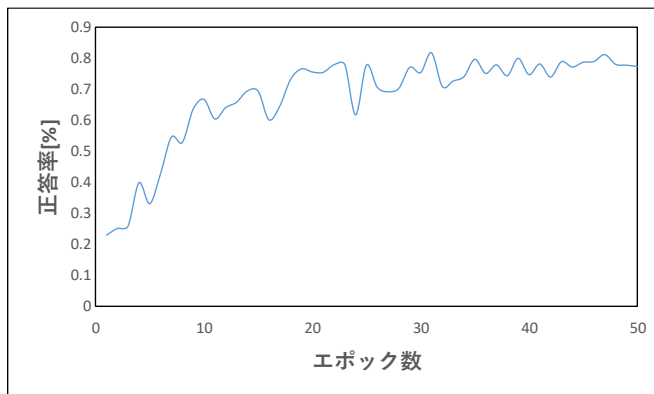


図 5 正答率のグラフ

6. まとめと今後の展望

RNN の推論・学習器を FPGA に実装することに成功した。また FPGA に実装した回路を用いて実際に音声識別を行い評価することで、実際に FPGA に実装できる程度の回路規模であること、設計した回路が有効であることを確認した。

今後の展望としては、まず最高正答率が 82%と決して高いとはいえない数値なので、隠れ層のユニット数や学習率などの決定に任意性があるパラメータの検討がある。ほかに、本稿では小規模な FPGA を用いて評価を行ったので計算速度の評価を行っていないため、大規模な FPGA ボード上に回路を実装し、計算速度の評価を行っていきたいと考えている。

参考文献

- [1] P. Werbos, "Backpropagation through time: What it does and how to do it", Proc. IEEE, **78**, 10, 1550-1560 (1990)
- [2] R. Williams, D. Zipser, "A learning algorithm for continually running fully recurrent neural networks", Neural Comput., **1**, 270-280 (1989)