

工学系研究科電気系工学専攻博士演習

Oakforest-PACS を用いた深層学習

埜 敏博

東京大学情報基盤センター

本稿は 2019 年度 9～10 月にかけて実施された、大学院工学系研究科電気系工学専攻博士演習の実施内容について紹介する。博士演習は同専攻の博士課程の学生が二人 1 組となって、2 ヶ月の間に、各教員が提供する課題に基づき演習を行うもので、原則として受講者の専門と異なる分野の課題を選択することとなっている。

今回の受講者は、それぞれ、高電圧工学、マルチメディア情報処理を専門にしている学生であった。このように大きく異なる分野でも今後共通に使えるであろう技術として、深層学習をスパコン上で試してもらうこととし、以下のようなステップで課題を設定した。

1. アカウント付き並列プログラミング講習会「MPI 基礎」の内容で Oakforest-PACS の利用方法に慣れる。
2. ChainerMN の環境を構築し、MPI を用いた複数ノードでのデータ並列による学習を実現する。
3. 画像分類の CIFAR100 データセットを用いて、複数ノードでデータ並列による学習を実行する。プロセス数、スレッド数などの組み合わせを変えて、性能について議論する。
4. (オプション) CIFAR 以外の異なるタイプの深層学習でも試してみる。

実際には 11 月まで延長し、2 名ともにオプションまで含めて実施できた。一人は、くずし字の平仮名の識字を行う Kuzushiji-49 の学習、もう一人は DCGAN (Deep Convolutional Generative Adversarial Networks, GAN:敵対的生成ネットワーク)によるアニメ顔画像生成に挑戦した。前者はデータセットが多く、クラスが 49 文字であるため、CIFAR100 よりもむしろ高い汎化性能を得られた。後者は、十分な学習が行われず、データ並列学習にはもう一工夫必要そうである。

今回は、教育利用の 16 ノード (68 コア×16 ノード=1088 コア)、15 分のジョブキューで実施したが、chainer のチェックポイント&リスタート機能を使って何度も再実行を繰り返すことで、十分な学習を実現できている。

博士学生であったことや、すでにスパコン利用経験があった学生もおり、当初想像したよりもトラブルが少なく実施することができた。

参 考 文 献

[1] Chainer, <https://chainer.org>

[2] The CIFAR100 Dataset, <https://www.cs.toronto.edu/~kriz/cifar.html>