

第 111 回お試しアカウント付き並列プログラミング講習会

「第 1 回 ディープラーニング分散学習ハッカソン」

下川辺 隆史

東京大学情報基盤センター

2019 年 1 月 24 日（木）、25 日（金）の 2 日間にわたり、東京大学情報基盤センター4 階 413 遠隔会議室にて、第 111 回お試しアカウント付き並列プログラミング講習会「第 1 回 ディープラーニング分散学習ハッカソン」が開催されました。

本講習会では、初めての試みとして、スパコンを利用したディープラーニング分散学習をハッカソンとして開催しました。ハッカソンは参加者がディープラーニングのモデルおよびデータセットを持ち込み、マルチ GPU、マルチノードでの分散学習を実践する形で進行します。本ハッカソンでは、すでに単一 GPU あるいは単一ノードにおいて複数 GPU を用いたディープラーニングを実践されている方またはチーム（3 人まで）を対象に、情報基盤センターに設置されたスーパーコンピュータ Reedbush (Reedbush-H) を活用した分散学習の実践を行います。情報基盤センター教員に加えて、Chainer, TensorFlow のスペシャリスト、及びエヌビディアのエンジニアがチューターとして参加し、お手持ちのモデルの精度や学習速度の向上をお手伝いします。

本講習会は、東京大学情報基盤センター、株式会社 Preferred Networks, TensorFlow User Group, エヌビディア合同会社, PC クラスタコンソーシアム（実用アプリケーション部会）の共催で開催され、下記の皆様にチューターとしてご協力いただきました（敬称略）。

- 福田圭佑（株式会社 Preferred Networks エンジニア）
- 鈴木脩司（株式会社 Preferred Networks リサーチャ）
- 藤原秀平（株式会社メルペイ / Google Developers Expert (ML)）
- 太田満久（株式会社ブレインパッド / Google Developers Expert (ML)）
- 成瀬 彰（エヌビディア合同会社 シニアデベロッパーテクノロジーエンジニア）
- 山崎和博（エヌビディア合同会社 ディープラーニングソリューションアーキテクト）
- 村上真奈（エヌビディア合同会社 シニアソリューションアーキテクト）

本講習会のスケジュールは表 1 の通りです。講習会の内容の詳細や講習会で使用した資料は、講習会の Web ページ¹に掲載しておりますので、そちらをご覧ください。受講者には実習で使用した Reedbush-H を受講後 1 ヶ月間利用できるお試しアカウントが与えられます。

今回の講習会では、応募のあったチームの中から 8 チームを選定し、そのうち 6 チーム合計 11 名が参加しました。受講者の内訳は、学部学生：1 名、大学院（修士）学生：5 名、大学院（博士）学生：4 名、その他大学教職員：1 名でした。講習会終了後に実施したアンケートの質問項目と回答の人数分布は表 2 の通りです。自由回答では、下記の回答をいただきました。

- 大変有意義な回でした。ありがとうございます。
- メンターさんに色々教えて頂き非常に勉強になりました。ありがとうございます。
- 知識の豊富なメンターのみなさまと直接お話できて大変勉強になりました。ありがとうございます。

¹ <https://www.cc.u-tokyo.ac.jp/events/lectures/111/>

ざいました。

- エラーが出た際に詳しい人に話を聞くことができとても良かったです。自分では分からないところが多くあるので、「これはこういう部分でのエラーだと思う」というアドバイスがとても助かりました。
- Chainer の開発に関わっている方にアドバイスをいただけて、得るものが大きく楽しかったです。cupy が速度のために敢えて numpy の全ての記法に対応していない話が印象的でした。
- ハッカソン参加者の為の slack チャンネルがあると良いかなと思いました。色々な情報をシェア出来るので
- バグや問題を共有できれば（ホワイトボードに書くか Slack を使うか）、他の人が分かるかもしれないので、助け合えると思う。
- 単一ノードでの GPU 並列は実現でき、大変勉強になりました。
- ハッカソンの趣旨がよくわからず、モデル作りから初めたため時間を有効に使えなかったアンケート集計結果を見ると、多くの方に満足してもらえたようで、改善点を踏まえて、次回以降の開催を検討したいと考えております。今回は1月という学生にとっては参加しにくい時期となってしまいました。アンケートでも6月から8月頃に開催して欲しいというご意見もありましたので、開催時期の変更も検討してまいります。

表1 スケジュール

日付	時間	内容
1月24日	09:30 - 10:00	受付
	10:00 - 10:30	Reedbush-Hの紹介とログイン
	10:30 - 11:00	分散学習基礎講座
	11:00 - 11:20	ChainerMN 紹介
	11:20 - 11:40	TensorFlowでの分散学習紹介
		(休憩)
	13:00 - 13:45	参加者・チーム紹介
	13:45 - 17:15	実践開始 (適宜自由に休憩)
	17:15	終了
1月25日	10:00 - 16:00	実践開始 (適宜自由に休憩)
	16:00 - 17:15	実践内容の発表と共有
	17:15	アンケート記入・終了

表2 アンケート集計結果の人数分布と平均

	評点	1	2	3	4	5	平均
講習会時間	短い <-> 長い	1	1	7	2		2.9
講習会講義内容 (プレゼン)	簡単 <-> 難			9	2		3.2
配布資料内容	簡単 <-> 難		2	6	3		3.1
満足度	不満 <-> 満足	1	1		3	6	4.1