

2018 年度 大規模計算論 授業概要

2018 High Performance Computing Class Overview

松岡 聡 Satoshi Matsuoka

2018 年 12 月 5 日(水) Wednesday, Dec 5, 2018

■講義の概要とねらい Course description and aims

本年度は、ハイパフォーマンスコンピューティング(HPC)・並列処理・クラスタ計算・クラウド・の分野において、エクサスケールに向かうスパコンや大規模 IDC にとって非常に重要な「HPC 的手法によるビッグデータ・AI の大幅な性能やスケールの向上」の学術的な成果の近年の論文をサーベイする。

This year's class will deal with "Immense acceleration of Big Data and AI through convergence with HPC - Scalable algorithms, systems and applications."

■授業の進め方 Class flow

初回のガイダンス時にハイパフォーマンスコンピューティング(HPC)・並列処理・クラスタ計算・IDC・ビッグデータ・クラウドの分野において「HPC 的手法によるビッグデータ・AI の大幅な性能やスケールの向上」にまつわる特定のテーマの英語の論文のリストを提示する。履修者はそれらのうち 1 本を選択し、**第二回目以降の講義でスライドによる発表を 1 回以上行う。**

- 発表中は履修者同士や教員・TA の間で内容に関する議論を行う。
- 発表時間は履修者の人数により変動するが、45~90 分程度を想定する。
- 発表に使用するスライドは授業当日までにメーリングリスト (hpc18@smg.is.titech.ac.jp) 経由で教員・TA・受講者に共有すること。

また、講義終了後に選択した論文とさらに選択した論文 2 本の合計 3 本に関するレポートを提出する。

In the first orientation, we will offer a list of English papers about “immense acceleration of Big Data and AI through convergence with HPC - Scalable algorithms, systems and applications.”

Each student will pick one of the papers, and **give a presentation about the paper one or more times.**

- There will be many questions that will be asked by the instructor, the TA, and the student audience. Questions will be asked to audiences as well from the instructor, and the TA, so full attention should be paid during the class.
- Each presentation will be 45 minutes to 90 minutes, varying with the number of students.
- Presentation will be done online using PowerPoint. A PDF version should be prepared and sent to the mailing list hpc18@smg.is.titech.ac.jp A DAY PRIOR TO THE CLASS so that the student audience can refer to it during the presentation. The presentation will also be archived at the following page.
 - <http://matsu-www.is.titech.ac.jp/lecture/lecture-wiki/index.php?hpc2018>

Students also have to submit a report about three papers they chose from the list, including the paper they presented.

■成績評価の基準及び方法 Student learning outcomes

成績は発表 (50%)、レポート (50%) により総合的に判断する。また、発表中に質問等を行った学生には講義ごとに 5%を加点する。

- The grades will be marked in the following fashion
 - Presentation: 50%
 - Report: 50%

- For each class, if a student voluntarily asks a question or states a technical opinion, additional 5% credit will be added

■授業日程・場所 Course schedule

- 12月5日(水)(ガイダンス): 3-4限(10:45-12:15)、5-6限(13:20-14:50) @ H116
- 12月18日(火): 1-4限(9:00-12:15)、5-8限(13:20-16:35) @ S518
- 12月19日(水): 1-4限(9:00-12:15) @ S518
- 12月20日(木): 1-4限(9:00-12:15)、5-8限(13:20-16:35) @ H117
- 1月7日(月)(予備日): 1-4限(9:00-12:15) @ W832

- December 5 (Wed) (Orientation): Period 3-4 (10:45-12:15), 5-6 (13:20-14:50) @ H116
- December 18 (Tue): Period 1-4 (9:00-12:15), 5-8 (13:20-16:35) @ S518
- December 19 (Wed): Period 1-4 (9:00-12:15) @ S518
- December 20 (Thu): Period 1-4 (9:00-12:15), 5-8 (13:20-16:35) @ H117
- January 7 (Mon) (Extra time slot): Period 1-4 (9:00-12:15) @ W832

■補足 Supplementary description

- ビッグデータと HPC の融合が近年様々な形で模索されている。これは HPC にまつわるデータ処理をビッグデータ技術を用いて大規模化・広帯域化・高速化するだけでなく、ビッグデータに固有な処理を HPC 技術を用いて高速化・大規模化する課題もある。
Convergence of HPC and Big Data are being explored in various ways. This includes acceleration of large-scale data in HPC, but also utilizing HPC technologies to accelerate big data processing such as graph processing and deep learning.
 - テーマの例 Example Themes:
 - ◇ Architectures for accelerating the handling of extremely large-scale data in HPC and AI.
 - ◇ Performance Modeling of Big Data and AI processing such as Deep Learning
 - ◇ Large Scale Parallel frameworks for Pre- and Post- processing of large-scale simulations using Big Data and AI methods
 - ◇ Utilization of new non-volatile storage and memory for large-scale Big Data and AI scaling
 - ◇ Acceleration of Large-scale Graph Algorithms on Supercomputing Platforms
 - ◇ Accelerating Deep Learning using GPUs and other HPC accelerators
 - ◇ Large-scale parallel filesystems and their algorithms
 - ◇ Utilization of Many-core processors such as GPUs and Xeon Phi for big data and AI
 - ◇ HPC-class networks for Big Data and AI.
 - We will not deal with small-scale “big data”, pure HPC, or purely theoretical big data issues such as advanced theory of deep learning.
- Although one paper is to be presented in a principal fashion, all three selected papers should be read and introduced as related topics. References to all papers should be indicated.
- Each class will involve presentation by one speaker; there will be many questions that will be asked by the instructor, the TA, and the student audience. Questions will be asked to audiences as well from the instructor and the TA, so full attention should be paid during the class.
- The report should summarize the general topic covering and including ALL THREE PAPERS regarding the state of the art in HPC and Big Data convergence