SG11
理学におけるデータ科学実践
機械学習で自然科学を読み解けるか

中野 直人

国際高等教育院附属
データ科学イノベーション教育研究センター　特定講師
理学研究科　連携講師

2018年度 MACS成果報告会
2019年2月7日

理学におけるデータ科学実践：機械学習で自然科学を読み解けるか

参加教員：　中野直人
余田成男（地球惑星科学専攻）
田口 聡（地球惑星科学専攻），宮崎真一（地球惑星科学専攻）

実験や観測からデータを取得してそれを解析することは，自然科学のいずれの分野においても普遍的に必要とされることである．昨今は機械学習などのデータ科学的手法が発達し，計算機能力の向上に伴い，様々な目的でデータが潜在的に持つ情報の抽出が試みられている．しかし，実際の理学研究においてもこれらの手法はどこまで有効なのであろうか．本SGでは，データ科学的手法を簡単なものから習得して，データ科学を身近なものにすることを目的とする．さらに，既存手法との比較による再確認，既存手法では得られない新たな理解，ここで学ぶ新しい手法ではできないこと，手法の数学的構造，などを整理することで，各手法の本質の理解を目指す．

具体的なSGの進め方としては，まず各手法の大まかなサーベイを行ってデータ科学を俯瞰した後，サンプルデータで手法に慣れるようにする．その後，数名ごとの班に別れて機械学習的手法を実践する．各自が実習課題を持つこととし，月に1・2回のペースで結果を持ち寄り全体で議論をおこなう．数学的構造について検討し，手法の改善や対象範囲を理論的に整理する班も設置可能とする．解析に用いるデータは，各自の研究に関わるデータを持ち寄るか，興味のある分野のデータの提供を受けるなどする．本SGでは機械学習の手法に主眼をおくため，RやPythonを用いて実習を行うが，初めのサーベイ時にチュートリアルを行うなど，計算環境の準備も行う．専門分野毎に得意とする理論や手法があるため，本SGを介して専攻横断的に方法論の共有も行えると尚良い．

理学におけるデータ科学実践：機械学習で自然科学を読み解けるか

参加教員：　中野直人
余田成男（地球惑星科学専攻）
田口 聡（地球惑星科学専攻），宮崎真一（地球惑星科学専攻）

概要

* 機械学習の手法を勉強をする
* 何でも良いから自分の興味ある問題に取り組む

やったこと

* 興味あるデータを持ち寄ったり，提供してもらったりしてデータ解析する
* ツールと手法のチュートリアル → 実習

  ~ Python + Jupyter Notebook

  ~ scikit-learn, TensorFlow (機械学習ライブラリ)

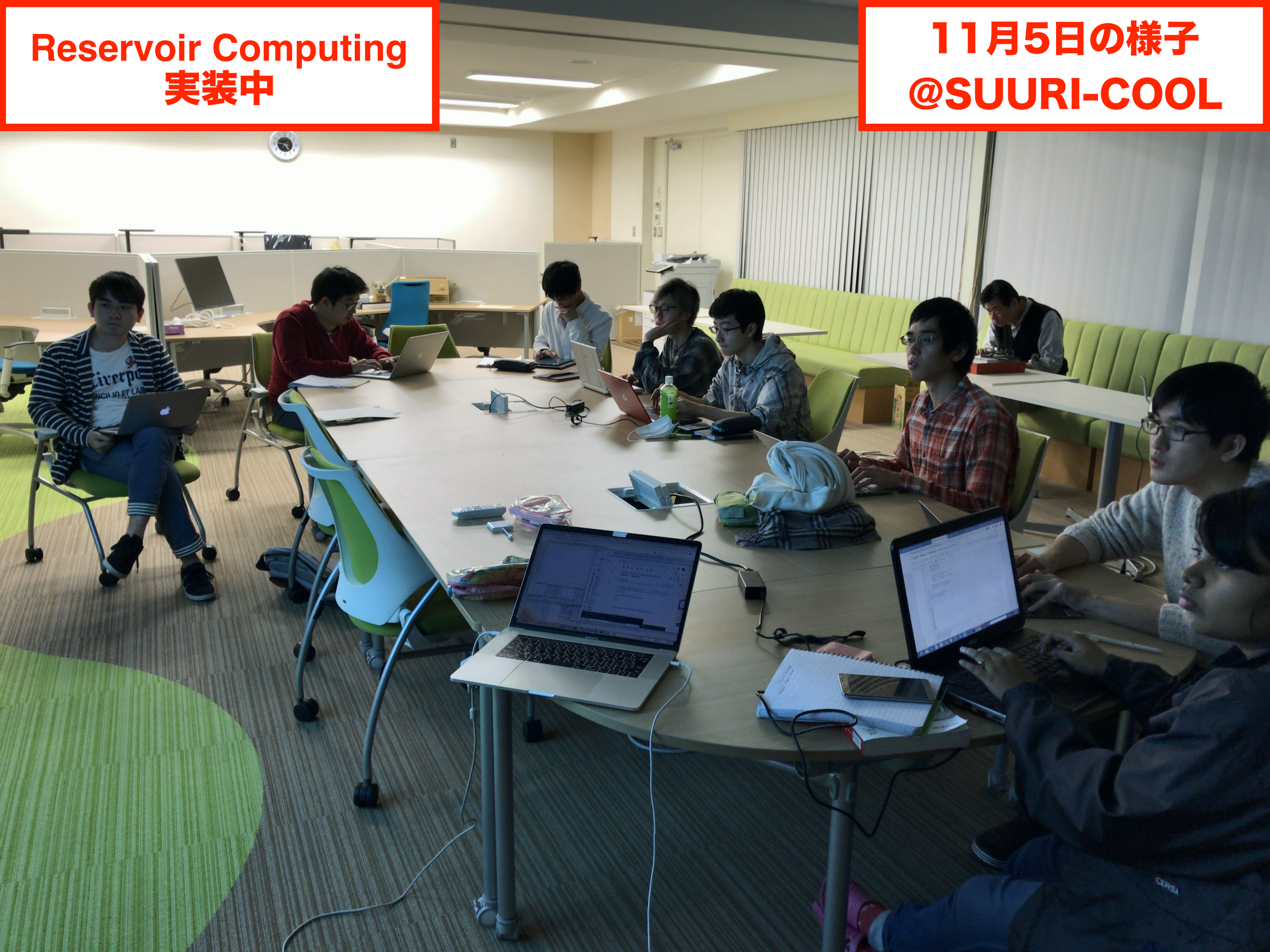
  とりあえず慣れる
* 各自の問題を（勝手に）やってもらう

登録者内訳

* 教員：中野，余田
  数学・数理解析専攻1名（+3名）
  地球惑星科学専攻3名
  物理学・宇宙物理学専攻1名
* 研究生1名
* 博士課程5名（D1 4名，D2 1名）
* 修士課程8名（M1 6名，M2 2名）
* 学部生5名（4回生3名，3回生1名，2回生1名）
  物理，地質，地震，数論，代数幾何，応用数学
* TA を3名雇用

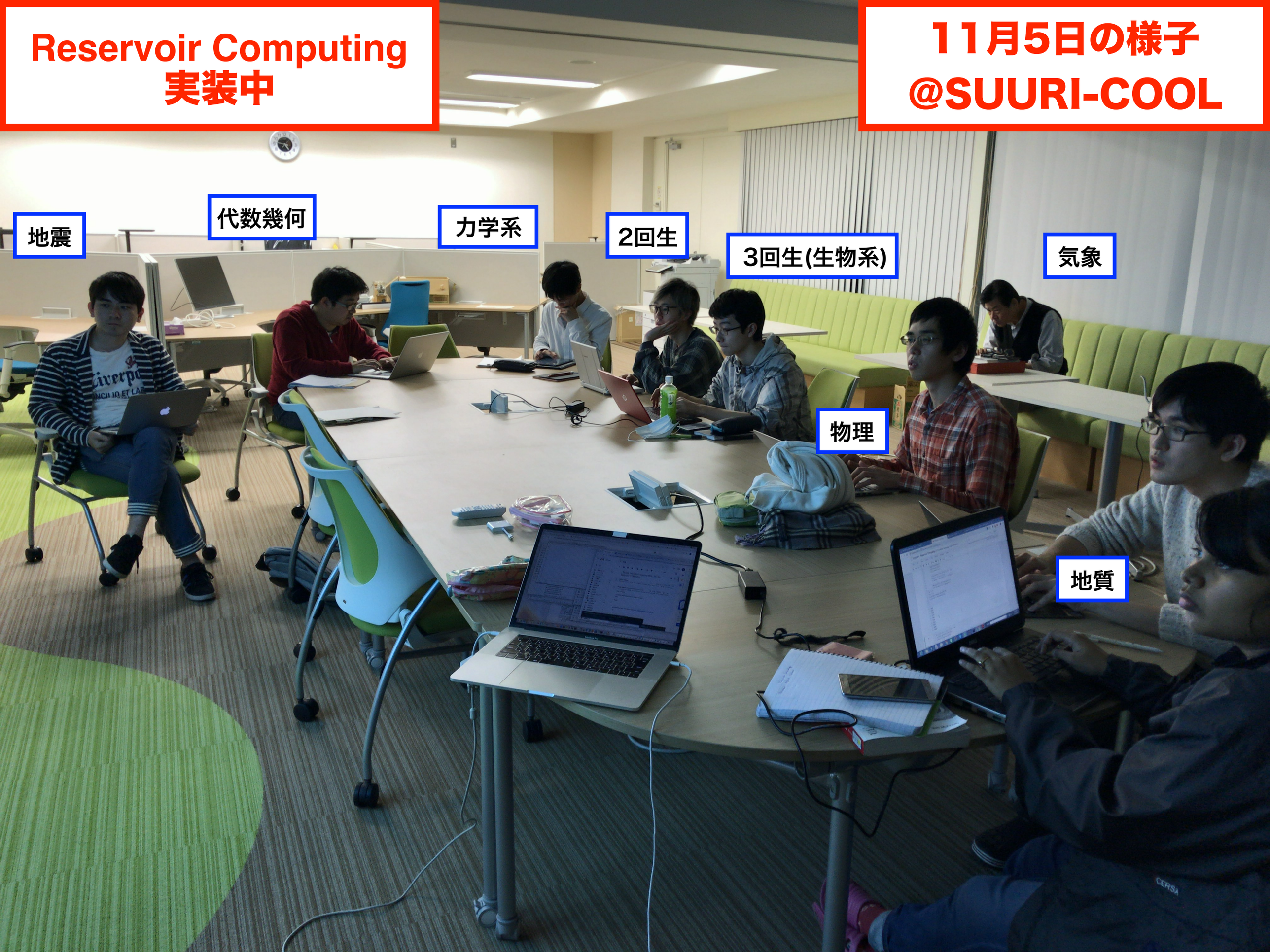
Reservoir Computing 実装中

11月5日の様子 @SUURI-COOL

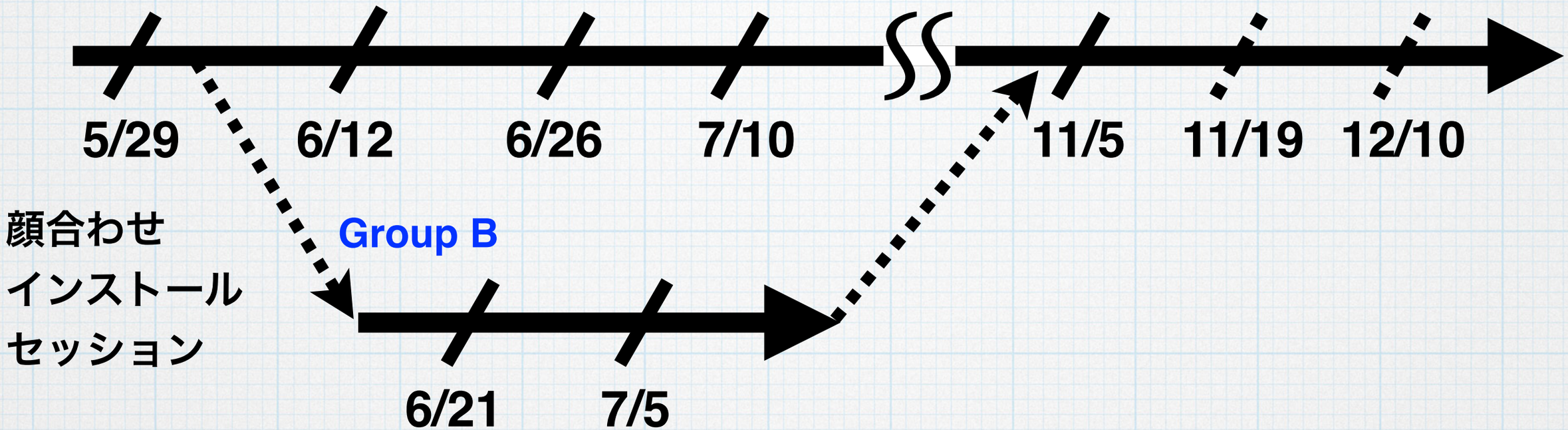


Reservoir Computing 実装中

11月5日の様子 @SUURI-COOL



これまでの活動



とにかくお手習い

各自の問題へ

実装だけでなく
関連する理論も

多層パーセプトロン
決定木・ランダムフォレスト
AutoEncoder
Reservoir Computing
畳み込みニューラルネットワーク

お手習いの内容

* 多層パーセプトロン
* 決定木・ランダムフォレスト
* AutoEncoder で次元縮約
* Reservoir Computing で力学系の再現
* 強化学習・方策勾配法で迷路を解く
* 畳み込みニューラルネットワークで画像解析

参加者の各自の問題

* Kaggle の laboratory earthquake の予測
* 決定木・MLP の層の深さの関連
* 画像解析手法を用いて統計物理の論文の追試
  * 第一励起状態波動関数の確率密度分布
* JRA-55（気象庁55年長期再解析）