

# MERIT 企業インターンシップ（国内）報告書

理学系研究科 物理学専攻

MERIT 5 期生

足立大樹

## 【実施期間】

2018 年 11 月 1 日～2018 年 12 月 27 日

## 【受け入れ先】

日本電気株式会社 データサイエンス研究所

## 【概要】

計算物質科学人材育成コンソーシアム (PCoMS) の支援の下, MERIT の了承を得て, 神奈川県川崎市にある日本電気株式会社 (NEC) データサイエンス研究所においてインターンシップを行わせて頂いた. 本インターンシップの目的は「深層学習を用いた画像処理技術の開発」であり, 主に近年画像処理系の機械学習の分野において盛んに研究が行われている「半教師あり学習」を行う深層学習の手法について, 性能評価や性能向上のためのアプローチについて研究を行った.

## 【活動内容】

画像のクラス分類問題は機械学習において非常に研究されている問題であり, 例えば MNIST と呼ばれる 0~9 の手書きの数字のデータセットの画像を 0~9 の 10 クラスに分類する問題は, 画像のクラス分類のベンチマークとして非常に有名である. しかしながら, 通常行われる画像のクラス分類は以下のような実用上の問題点が存在する.

1. 学習データセットのラベルの割り振りを全て人間が行う必要がある
2. 非常に偏ったデータセット (特定の数字が極端に少ない等) の場合, クラス分類の正答率が下がる

これらの問題を解決するために様々な方法が考案されているが, 本インターンにおいては特に半教師あり学習の手法の一つである Domain Separation Network (DSN) と呼ばれる深層学習のネットワークを用いた研究を行った.

半教師あり学習とは, あるラベル付けされたドメイン (ソースデータセット) を用いて別のラベル付けされていないドメイン (ターゲットデータセット) のクラスを推定するための方法である. 例えば, ある人の手書き数字データセットがラベル付けされている場合, それに最適化した分類機を用いて別の人の手書き数字を分類しようとしても, クラス分類の成功率はそれほど高くはならない. しかし, 「ラベル付けされたデータセットとラベル付けされていないデータセットが異なるデータセットである」という情報を加え, 2つのデータセットを同時に用いて学習させた分類機を用いると, ラベル付けされていないクラスに対するクラス分類の成功率を大きく上昇させることができる. これはクラス分類を社会の問題に適用する上で非常に重要なテーマであり, 例えば, 様々な方向から 3次元の CG データを撮影したものを学習データとする場合など, ラベル付きの画像データを機械的に生成することが容易な場合には, 人がデータセットをラベル付けする手間を簡略化することができる.

今回のインターンシップでは、既に出来上がった DSN 用のソースコードを実際の様々なデータセットに対して用い、性能評価や性能向上のためのネットワークの改良等を行った。本インターンシップで得られた知見は、本研究テーマの目的である画像処理のみならず、様々な分野においてクラス分類を行う際、もしくはクラス分類を行うためのネットワークを考える際に、非常に重要なものであると考えられる。

#### 【謝辞】

本インターンシップの機会をくださった日本電気株式会社様に深く御礼申し上げます。特に、受入研究者として事務手続きや研究に関してご指導をして頂いた柴田様、研究で得られた結果についての議論、質問から通信ネットワークの問題に至るまでの幅広い内容に対する回答を懇切丁寧にくださった並木様、今回の研究内容を通じて企業において研究活動を行う際の指針を教えてくださいくださった谷内田様を始め、データサイエンス研究所の皆様には大変お世話になりました。また、機械学習を扱う研究室からのインターン生として同じく研究を行ってまいりました安藤様、平良様には、機械学習全般における内容について議論して頂きました。受け入れ先企業との調整等に尽力してくださった有馬様はじめ PCoMS の皆様にも大変感謝しております。最後になりますが、本インターンシップの実施を快諾してくださった指導教官の藤堂教授と副指導教官の田中教授、そしてこのような機会を与えてくださった MERIT プログラムに感謝申し上げます。