

論文の締め切りなど色々あって、提出大分遅れてしまいました。今週からもう学期が始まっているのですが、先学期のことと冬休みやったことを書きます。前回書いた時に比べて、論文が去年あった機械学習でもっとものいい学会の一つの Neurips に 2 本通りました。

[1] Doubly Robust Off-Policy Value and Gradient Estimation for Deterministic Policies

[2] Off-Policy Evaluation and Learning for External Validity under a Covariate Shift

という論文です。[2]の話は Spotlight(普通のアクセプトだとポスター発表なので、もっと選ばれている)に通って日経の方に取材してもらいました。その方によると、日本人の Spotlight や Oral がかなり増えているらしいです。とてもいいことだと思います。

あと 8 月あたりに Journal of machine learning research という機械学習のジャーナルにも論文が通りました。機械学習の研究者を目指すのであれば、ジャーナルに出す必要はないというのが一般論だと思います。ただ、将来的にアプライするかもしれない、統計、OR、ビジネススクール etc のアカデミックポジションだとそれらの分野のジャーナルがある程度なければあまりとってくれないらしいので、ジャーナルにも並行してもっと出していくようにしていきたいです。

先学期(秋学期)は授業かなりとって、研究の時間が大分減りました。あとは受けた授業と最近、書いた論文について紹介します。

## 研究

### Finite Sample Analysis of Minimax Offline Reinforcement Learning: Completeness, Fast Rates and First-Order Efficiency

<https://arxiv.org/abs/2102.02981>

オフライン強化学習において、Neural Network やカーネルなど様々な関数近似手法に基づいた Minimax 手法の(非漸近論の意味での)収束レートがどのように変わるか示した論文です。UIUC にいた時のアドバイザーと友達、コーネルのアドバイザー達、東大の今泉さんとの論文です。貢献は(1) Q 関数とその双対関係にある W 関数の Fast Rate(回帰問題の意味での)を達成できることを示した、(2)オフラインのポリシー評価の First Order Optimality が達成できることを示したことです。今までの自分の論文は漸近評価が多くて、もっと現代的な非漸近理論をしっかりと使った論文を書いたことはなかったのですが、それらを使うために書いている中で色々学べてためになりました。あと様々な共著者と議論できて、楽しかったです。内容的にも特に自信作と思っているので、うまく評価されればいいなと思います。

### Fast Rates for the Regret of Offline Reinforcement Learning

<https://arxiv.org/abs/2102.00479>

Q-Learning が Fast Rate(分類問題の意味での)より早いレートを得ることができることを示した論文です。コーネルのアドバイザーと友達との論文です。学習理論で Tsybakov による有名な結果として、一般的な普通の分類問題を考えたときノイズにうまい条件があれば Fast rate を達成できるという事実があります。この論文はその強化学習バージョンです。

### Optimal Off-Policy Evaluation from Multiple Logging Policies

<https://arxiv.org/abs/2010.11002>

複数の種類の過去データがある時に、そこからどうやって、因果効果を推定するかという論文です。昨年、同じ問題を考えた時に、推定下限をどう出すのかよくわからなくて考えることをやめたのですが、他の Neurips に通った論文書いていた時に出し方がわかったので、論文にしました。コーネルのアドバイザーと船井の奨学生になる斎藤さんとの論文です。

### 授業

#### Foundation of Reinforcement Learning

強化学習の理論の授業です。コーディングはなくて、証明がメインでした。最先端の強化学習の理論がうまい形でまとまっていて、とてもためになる授業でした。強化学習は理論ベースのいい本があまりないのですが、授業ノートが本になっていて、とてもわかりやすいので広めていけたらと思います。

#### Analysis of Algorithm

アルゴリズムの授業です。これもコーディングはなく、証明がメインです。マッチング、組み合わせ最適化、計算量(NP hardness など)を学びました。演習がパズルみたいな感じで、かなり難しかったです。TCS (theoretical CS) 分野で研究している人は頭良くて凄いなと思います。実際、TCS の人が機械学習に入ってきて、競争の激化にしっかり貢献している気がします。

#### Causal Inference

コロンビア大学の授業を聴講しました。Causal DAG が中心の因果推論の授業です。この授業は Causal DAG の基礎を作った有名な Pearl の弟子の授業でした。ハーバードで Causal DAG の授業を受けた時と比べて、かなり視点が違っていました。特に様々な条件下での統一的な Identification の仕方を学べたのが良かったです。因果推論の不幸な歴史上、グラフ(Dag)を使った因果推論に意味あるかないかみたいな議論があり、Harvard の統計学科の有名な教授(Rubin)が特にいらぬという主張をしていて自分もそうかなと思っていたのですが、学んだ上だと個人的には理論的に面白いし、実用上での意味でも必要な分野だなと思います。おそらく、Causal DAG の理論における ID algorithm など最も重要な結果をわかりやすく説明した本がなくて食わず嫌いされている気がします。

### **Selected Topics in Applied Probability**

Optional Stopping の授業です。Optional Stopping の分野だと結婚問題とかが有名な例があります。結婚問題とは  $n$  人と順番にお見合いして、会ったらすぐ結婚するか決めなければいけないかつ一度、断ったら二度と結婚できないという状況で、一番タイプな人と結婚できる確率を最大にする戦略を考えるみたいな問題です。他にも人を雇うみたいな状況で同じ問題が出てきます。応用確率論のテクニックが必要な分野で、これもまた研究レベルに持っていくのは難しいなと感じました。

### **Bandit のセミナー**

授業ではないけど、夏から毎週 or 隔週ぐらいのスピードで Bandit algorithms という本を輪読しています。財団生の古賀さんとも輪読していてためになっています。研究でいつか使えればいいかなと思います。

### **最後に**

オンライン強化学習の研究していたのですが、頓挫してしまったので、今年こそは何かまずその分野で結果が出せたらなと思います。オンライン強化学習は早いペースで質もいい論文が出ていて分野としてなかなか大変です。あと、グラフベースの因果推論 (Causal Dag の知識) もたまってきたので、それでも何かいい結果が出たらと思います。