

船井情報科学財団報告書 2022/2/14 上原雅俊

最近締め切りが多くてかなり提出が遅くなってしまいました。最近アクセプトされた論文、インターン選考、学会、新しい論文について簡単に書きます。

最近アクセプトされた論文

前回の報告書の時と比べて First または Cofirst の何本か論文がアクセプトされました。

Uehara, Masatoshi, Xuezhou Zhang, and Wen Sun. Representation learning for online and offline rl in low-rank mdps. ICLR (Spotlight), 2022.

Uehara, Masatoshi and Wen Sun. Pessimistic model-based offline rl: Pac bounds and posterior sampling under partial coverage. ICLR, 2022.

Jonathan D Chang, Uehara, Masatoshi, Dhruv Sreenivas, Rahul Kidambi, and Wen Sun. Mitigating covariate shift in imitation learning via offline data without great coverage, Neurips 2021

Nathan Kallus and Uehara, Masatoshi. Efficiently breaking the curse of horizon: Double reinforcement learning in infinite-horizon processes. Operations research, 2021.

機械学習全般だと Neurips, ICML が一般的に昔からトップ会議となっていますが、ICLR は最近出来た Deep Learning 向けの学会でそれに並ぶような形になってきています。将来的に無くなっても不思議ではないと思うので、普段出さないのですが、ホットな学会ということで出してみました。また Operations Research は OR の中で一番いい雑誌です。とりあえず一つ OR のパブリケーションがもてて安心しました。ただ提出してから2年以上かかり、いいジャーナルに載せるためのプロセスは色々面倒だと学びました。

インターン選考

12月から1月は機械学習のパブリケーションをゴールとしているリサーチインターンの面接をかなり受けました。結局、MSRとAmazonになる予定です。せっかくなので、詳しく落ちたところの経験も踏まえてかきます。

まず12月ごろに Deep Mind の面接がありました。Deep Mind はまず一次のテクニカル面接（簡単な数学とかプログラミング）があって、二次からチームマッチングの面接が始まります。去年も受けていて、一次は去年のものが使いまわせたので、二次から始めました。強化学習のチームは何個かあるのですが、2次ではコンタクトをとったロンドンとパリの強化学習チームからの選考を受けました。5回面接があって、通るかなと思っていたのですが、最後に落ちました。実装経験が乏しい理論系の人のはかなり限られているなど理由は色々あると思いますが、自分の反省点として、ほとんどインタビューの準備をしなかったのが、不味かったなと思いました。

次に Meta AI のインタビューがありました。メンターの方は知らなかったのですが、強化学習、Bandit の理論系の人を求めている公募が twitter に上がっていたので応募しました。1次は Coding Interview で、2次はメンターとの面接でした。Coding Interview は本当に簡単で、基本特に選考に関係なかったと思います。2次は前回の反省からかなり準備しました。その回もあって、オファーをもらえました。

そのあとに Microsoft Research (New England) との面接がありました。メンターの方は知らなかったです。面接は1回でした。今回は公募情報からプロジェクトがはじめから大体わかったので、簡単なサーベイを書いてプレゼンしました。その熱意も伝わり、オファーをもらえました。

最後にアマゾンで強化学習、Bandit を使っているチームと面接しました。事前にメンターとはコンタクトをとって、他にオファーがあるので早くして欲しいと伝えて、面接が始まりました。早くするために1個おそらく面接を飛ばしたのですが、Coding + behavioral Question の面接と、メンターの面接が同じ日に2個ありました。残念ながら Coding は簡単なのに全然できませんでした。それでも、メンターの人と面接した後、数時間後にオファー貰えました。

他の会社や、同じ会社の他のチームも受けていたのですが、これ以上書くときりがないので、まとめると以下が個人で出来る範囲内で後悔しないことかなと思います。1) コンタクトをたくさんとる 2) 自分の今やっている研究、将来やりたいことを短く的確にいえるように用意しておく 3) インタビュー側が研究者の場合、その人の論文読んでおく 4) 事前にインターンのプロジェクトがなんとなく分かっている場合、自分のアイデアを詳細に言えるよう準備しておく（情報があまりなくてもないなりに何やりたいか準備しておく） 5) 典型的な Behavioral Question にも準備しておく。（Deep Mind とアマゾンだと30分ぐらいこれだけのための面接がありました）おそらく分野にもよると思いますが、リサーチ系の Coding interview は本当

にまずい人を落とすため程度にしか使っていないと思うので、重要ではない気がします。
Coding 能力を見たくても、リサーチから判断されると思われます。

学会

学会は INFORMS という OR で一番大きい学会と、Neurips に参加しました。Informs はアナハイムであって場所は楽しかったですが、正直、発表はあまり面白いものがなかったです。Neurips はオンラインでした。今回は因果推論のワークショップのオーガナイザーをやりました。

<https://sites.google.com/view/causal-sequential-decisions/home>

レビューを書いてくれる査読者をあつめたり、宣伝したり、スピーカーを集めたり大変でしたが、今度からはもっとストレスなくやれるようになったかなと感じます。

論文

前回から 3 本論文を書きました。

Representation learning for online and offline rl in low-rank mdps.

<https://arxiv.org/abs/2110.04652>

コーネルのアドバイザーのプリンストンのポスドクとの論文です。どのように良い Representation を学びながらオンライン強化学習を行うべきか示した論文です。よくある Linear Bandit のモデルだと表現 (Feature) が既知という仮定が課されていますが、この論文では表現が未知という難しい問題を考えています。提案アルゴリズムは単純で、表現を最尤法で学んで、それに基づいて UCB を行うというものです。既存手法に比べて Sample Complexity が優れていて、計算的にもシンプルで優れているが特徴です。今までの手法は理論保証 (PAC guarantee) をえるために、アルゴリズムがかなり複雑になっていました。私たちの手法はシンプルなのに正しい理論保証とよい Sample complexity があり、強化学習の理論コミュニティにはかなり驚くべき結果が与えられたと思っています。結果的にも ICLR のスポットライトに選ばれました。

A minimax learning approach to off-policy evaluation in partially observable markov decision processes

<https://arxiv.org/abs/2111.06784>

London of School Economics と UIUC のコラボレーターとの論文です。強化学習だと未測定
の交絡因子がないという仮定をよくおきます。この論文では未測定
の交絡因子がある場合に、どのように任意の関数近似（ニューラル
ネットなど）を組み込みながら学習するか Confounded POMDP の
枠組みで定式化して示した論文です。テクニカルには 2 種類の操作
変数をうまく使うかというのが肝になってきます。前書いた論文
(Causal inference under unmeasured confounding with negative
controls: A minimax learning approach) の強化学習バージョンにも
なっています。

**Efficient reinforcement learning in block mdps: A model-free
representation learning approach**
<https://arxiv.org/abs/2202.00063>

最初の論文 (**Representation learning for online and offline rl in
low-rank mdps**) はモデルベースでした。この論文ではその論文を、
モデルフリーに拡張しました。