

留学報告書 (2024 年 12 月)

1. はじめに

大古 一聡と申します。UC Berkeley EECS の博士課程 1 年生です。専門は機械学習理論です。

2. 渡米

ビザの準備が大変でした。4月に留学先を決めたらすぐ取り掛かった方が良かったです。僕は過去にもアメリカのビザを取っていたので大使館で面接を受ける必要はなく、パスポートと書類を郵送するだけでしたが、それでも危なかったです (ESTA は前は良かったらしいのですが、今は面接を受けないといけません)。僕がビザの申請をしたのは7月でした。7月はCOLT という国際会議と ICML という国際会議があったので、正味2週間しか日本にいなかったです。7月7日の朝にエドモントンから帰ってきて、7月8日の朝にパスポートを大使館に投函しました。ブログを読み漁った感じだとその週の終わりには処理されてもおかしくなかったのですが (処理されたかはネットで確認できます)、結局確か7月16日に処理がされ、受け取ったのは18日だったような気がします。20日にはウィーンに行ったので、マジでギリギリでした。ほんと焦りました。

これ以外に準備らしい準備は出発前日の夜7時まで特にしていません。その夜は秋葉原のユニクロ (営業時間が長い) で服を買まくって、新しいスーツケースを買いました。それでも今日 (11月10日) までなんとかこなっています。

準備がいらなかったのは寮に入ったからというのが大きいです。食事と寝る場所が確保されているので、服だけ持ってきてもどうにかかりました。食事も美味しいので大変満足しています。周辺の家と比べると高いのですが、知り合いが「博士1年目はコストをかけてでも生活に慣れた方がいい」と言っていたのでそうしました。これは正解でした。

日本は寒そうですがカリフォルニアは暖かいです。正直天気が良い以外にカリフォルニアの良いところを知らない (そしてアメリカの天気が悪い場所は良いところを全く知らない) ので、逆に言うとカリフォルニアはシンプルに天気が良いというだけで発展した訳です。というカス理論を信じられるくらい天気が良いです(?)。今後も太陽の光を浴びて伸び伸びと研究をしていきたいと思っている夜型人間です。

3. 授業

すごくベーシックな機械学習の授業 (学部生も取るようなやつ) と、vision and language model の輪読の授業を取っています。前者を取ったのは英語でテストを受けることに不安があったからですが、中間試験の結果は悪くなかったので、来学期はもっと難しそうなクラスを取っても大丈夫な気がします。後者は合同研究室セミナーのような感じで、色々な知見

を得られて面白いです。先週発表を終えたのもう安泰です。

4. 研究

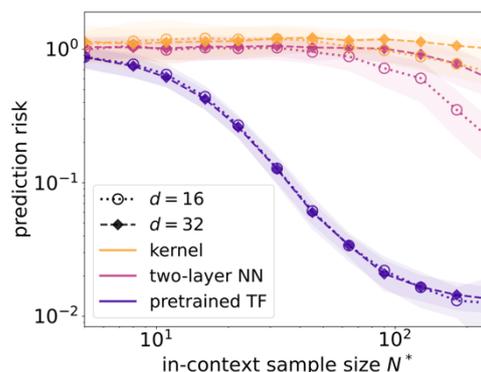
新しく初めたプロジェクトはどこまで内容を言って良いのか分からないので出てから紹介しますが、overleaf は 72 ページあってそれなりに良い出来です。

とは言えこの 2 ヶ月は日本から持ってきたプロジェクトにもそれなりに時間を取られていました。本当はスパッと終わらせて来れると良かったのですが、なかなかそうも行かないですね。完全にこっちのこただけ出来るようになるのは来年になってからだと思います。

最近出した論文で一番面白いと思うものを宣伝します。(これは東大で同級生の宋くんが頑張ってくれました。) ChatGPT に「 $(x,y)=(1,1)$ のとき $f(x,y)=1$, $(x,y)=(2,1)$ のとき $f(x,y)=2$, $(x,y)=(3,1)$ のとき $f(x,y)=3$ です。 $(x,y)=(4,1)$ のときは？」と聞くと $f(x,y)=4$ と答えてくれます。回答を生成する時、別にモデルのパラメータをアップデートしているわけではありませんが、与えられた文章中で示された入出力関係を元に、予測を行うことができます。通常学習と言ったらパラメータをアップデートすることですが、パラメータをアップデートしなくても文章中で示された入出力関係を「学習」できるというのは一つ最近の言語モデルの特徴的な能力な訳です。これを「文脈内学習」と呼びます。

大規模言語モデルの学習は、100 億円単位のお金をかけて、大量のデータ $((x,y)$ と $f(x,y))$ を見せる必要があります。しかし、先の例ではその必要はない。また、3 点が指定されても $f(x,y)$ は無数にあり得る訳ですが、その中で最もシンプルなものを選んでいる。それはなぜか？—それは、ChatGPT が僕の質問に答える前にあらかじめ、「1 変数にだけ線形に依存する関数がくるだろう」と決め打ちしているからです。

「大量のデータで決め打ちの仕方を学べば、少ない数の例示で文脈内学習ができる」これを理論的に示したのが今回の研究です。 d を変数の数とします (先の例では x と y なので 2)。そしてそのうち 8 つにだけ依存するような関数を学習することを考えます。 d が大きくなれば、当然、決め打ち無しでは学習するのが難しくなります。では横軸に文脈内で与えられる例示の数、縦軸に損失を取り、その関係を見てみましょう。黄色 (古典的手法) と赤 (Transformer ではないニューラルネットワーク) は、 $d=16$ の時と $d=32$ の時で損失の落ち方が違います (後者の方が遅い)。一方、大量のデータで学習された Transformer (GPT-2) は、 $d=16$ の時も $d=32$ の時も損失の落ち方が同じです。つまり、こういうことです：「問題の分布」に関する知識がなければ、古典的手法や Transformer ではないニューラルネットワークがそうであるように、変数の数が増えるほど問題は難しくなります。一方、 d がどれだけ大きくても、そのうち 8 つにだけ依存するということがわかってしまえば、そ



の 8 つだけ見れば良いわけで、与えられる変数の数が $d=16$ だろうが 32 だろうが、Transformer にとっては同じくらい簡単な問題だということです。

5. 最後に

Industry との距離が近いからこそ AI や ML は研究分野として盛り上がっているんだと思いますが、いざ身近な人が industry に行くとなると、それは一段違う話になります。「おお、そうか（もう一緒に論文書けなくなるのかな）」という一抹の寂しさと、「なるほどね（いつか自分にもそうやってキャリアを選ぶ時が来るんだろうな）」という淡い予感が入り混じって、現実を突きつけられたような気分になります。同クラが結婚したって聞いた時も、一瞬、人生について考えたりしましたが、気の迷いは 10 秒くらいで記憶の彼方に飛んで行きました。しかし今回ばかりはそう簡単に流せません。なんせこっちの方がよっぽど現実味がありますから。この業界は生き急いでる人だらけです。少なからず影響を受けてしまいます。最後になりますが、この場を借りて、留学を支援していただいている船井情報科学振興財団の皆様にご感謝申し上げます。