

第五回留学報告書

Funai Overseas Scholarship 2020 年度奨学生
古賀樹

2022 年 7 月

2020 年度から University of California San Diego の Computer Science 専攻 Ph.D. 課程に在籍している古賀樹と申します。Ph.D. 生活の 2 年目が終了し、この夏はインターンのためベイエリアに滞在しています。この報告書では 2 年目冬・春学期の研究や生活についてのご報告をさせていただきます。

1 研究

Ph.D. 一年目に取り組んでいた研究成果^{*1}が International Conference on Artificial Intelligence and Statistics (AISTATS) に無事通り、3 月末の学会で発表を行いました。内容としては（前回の報告書にも記載しましたが）、時系列データを差分プライバシー^{*2}を保証しながら公開する際に、時間軸における Subsampling を行うことで有用性を保ちながらより強いプライバシーの保証ができることを示したものです。

2 年目後半は二つのプロジェクトを主導しています。

一つ目は surveillance におけるプライバシーに関する研究です。例えば、ある時間に東京駅に人が何人いるか、といった情報に興味があるとします。しかし人がいつどこにいたかという情報はもちろんプライベートなもので、一般には公開されるべきではありません。東京駅にはたくさん人がいるので、一つの統計量だけから個人が東京駅にいたことが特定されることは現実には考えにくいですが、クエリの柔軟性が上がったり（e.g., サンディエゴの大学に通っ

ている人である時間に東京駅にいる人数）、そもそも人がほとんどいない場所におけるクエリ（e.g., 深夜 3 時のコンビニにいる人数）を想定すると統計量から個人が特定されてしまいます。そのため、そのような特定が不可能な形でノイズを乗せることで差分プライバシーを保証します。ところが差分プライバシーの定義に厳密に則ると、東京駅にいる人数に関するクエリを発行するたびに、全世界中の人にプライバシーロス^{*3}が一様に発生します。すなわち、その時間に東京駅に実際にいた人とジンバブエにいた人が同様にプライバシーの損失があったこととなります。これは完全に間違った考え方ではないのですが、直感的には現実との乖離があると言えるでしょう。私が行っているのは、このギャップを埋めるプライバシーの定義を考案する研究です。

二つ目は複数機関からデータを集める設定での因果推論におけるプライバシーの研究です。この研究は Duke 大学の先生とのコラボプロジェクトです。医療データを用いて因果推論を行いたいが、一つの医療機関からのデータを用いるとバイアスが生じる and/or データ数がそもそも少なく信頼度の高い推論が行えないという問題に対して、複数の機関からデータを集めて推論を行うことを想定し、その際に

^{*1} <https://proceedings.mlr.press/v151/koga22a.html>

^{*2} 現在アカデミアで盛んに研究がなされており、様々な企業でも採用されているプライバシーの定義

^{*3} ノイズを加えても完全にはプライバシーが守れるわけではないので、プライバシーロスを定義し、その程度を定量化する

どのようにプライバシーを保護するか、またプライバシー保護によりどのようなユニークな問題が発生するのかについて考えています。

二つのプロジェクトを同時に主導する経験が初めてで、どちらのプロジェクトも夏までには形にできず自分の力不足を少々感じた今年前半でした。しかしプロジェクト自体はとても面白く、コラボプロジェクトも非常に新鮮かつ刺激的なので、秋学期にはどちらも論文に仕上げられるよう精進します。

1.1 Research Exam

UC San Diego の CSE では卒業要件の一つに Research Exam というものがあります。研究分野のサーベイ論文を書き、指導教員を含まない教授三人からなるコミッティーの前でサーベイ内容と自分の研究について 1 時間程度のプレゼンをするという内容です。指導教員から落ちる人はほぼいないと聞いており、その通り終始和やかな雰囲気で行われました。フィードバックも良好でした。といっても準備にはある程度の時間がかかり、研究の進捗が阻害されたことは否めません。一方、プライバシーの研究に関するプレゼンを分野外の人にする非常に良い訓練になりました。

2 インターン

夏は Apple でインターンをしています。Apple はいち早く差分プライバシーを保証したプロダクトを世の中に出し始めていて、優れた研究者も在籍しているため、インターンが決まり非常に嬉しく思いました。内容としてはプライバシー x 機械学習の研究なので大まかには大学での研究と被りますが、大企業でしかできないような研究トピックに取り組むことができおり、非常に刺激的です。インターンの成果を論文にすることもできるので、次回以降の報告書で報告できるよう残り二ヶ月半努力します。

ベイエリアでは、会社負担でキッチン付きのホテルの部屋に住みながら、Apple Park という新しくで

きた本社のキャンパスで働いており、日々「最高だなあ」と感じています。特に午後、昼食後眠くなったからキャンパス内でジェラートを食べてリフレッシュし、仕事をした後はピックアップサッカーやジムでの運動で汗を流して、無料の社食で夕食をとり、大音量の音楽を流しながら運転して帰る、というルーティンがとても気に入っています。今後のキャリアを考える上での良い機会でもあるので、じっくり楽しみながら吟味していこうと思います。

3 講義

冬学期にのみ一つ講義を受講しました。これで卒業までは残り二つの講義を取るのみです。

3.1 CSE280A (Algorithms/Computational Biology)

所謂バイオインフォのコースでした。内容は担当教授の興味にかなり偏っていて、狭く深く、といった印象でした。DNA 配列における変異がどのように起き、それがどのように世代間で受け継がれていくのかについて統計・CS 的な手法で解析していく術を学びました。

4 生活

生活面に関しては特筆すべきこととして以下のようなことがありました。

- 家族が GW にサンディエゴに遊びにきてくれた
- NY に旅行に行った
- ベイエリアで船井同期とご飯を食べた
- 引き続き Intramural Sports*4 でサッカーをした

*4 サークル対抗戦のようなもの

