

船井情報科学振興財団第九回留学報告書

大谷直樹

2021年7月

2017年秋よりカーネギーメロン大学 (Carnegie Mellon University) の Language Technologies Institute (LTI) に在学しながら自然言語処理 (NLP) の研究をしている大谷直樹です。新型コロナウイルス感染症のワクチン接種も済み、とうとうキャンパスに戻る許可が降りました。が、この一年で完全に引きこもり体質になってしまったのでまだ戻る気になっていません。

1 近況報告

前回の報告書に書いたとおり、前学期の後半を日本で過ごしました。日本に居てもほとんど学業的な不便はなかったのですが、サマーインターンの手続きの事情で春学期が始まる直前にアメリカに帰ってきました。当然食生活のクオリティは爆下がりしましたが、無事に数ヶ月で慣れました。

渡米した1月当時日本では感染者数の増加がなかなか収まらない状況でしたが、成田からピッツバーグまでの移動は驚くほどノーマルでした。まだPCR検査の陰性証明が必要になる一週間前だったので¹、出国時はいつもどおりビザを確認されるだけで、シカゴでの入国時は検査・隔離や滞在先の確認すらなく、すぐに完全開放されました。国内線乗り換えも自由、公共交通機関の利用も自由でした。(場所や条件によっては状況が違ったのかもしれませんが...))

アメリカでは幸い COVID-19 に感染することもなく生活していたのですが、5月にベル麻痺という病

¹とはいえ一応成田で自主的に PCR 検査を受けてから行きました。



(a) ただいま、ピッツバーグ…。



(b) 4月、暖かくなったのでアウトドア席で食事。



(c) ワクチン接種記念。私はこの後に起こることをまだ知らない。

気にかかり、3週間ほど顔の半分が動かせなくなっていました。はっきりした原因は不明ですが、医者には発症の少し前に受けたワクチンとの関係を指摘されました²。顔半分の筋肉が動かせなくなると、笑い顔が何か企んでいる感じの表情になってしまったり、シャワー時に半開きの目の中に水が入ってきたり、唇に力が入らなくて飲み物を吹き出したりと、とにかく生活が地味に不便になるという学びを得ました。幸いにもこれといった苦痛はなかったのですが、外見上のインパクトが大きいせいで色々な人に心配され優しくされました。今回初めて数回通院するレベルで医者にかかりましたが、保険適用前で3,000ドルくらいの料金が発生していました。アメリカ、どうなってんの…。

2 学業

前学期に引き続き LTI の招待講演シリーズ『LTI Colloquium』³のティーチングアシスタント (TA) をしました。TA と言いながら事務手続きがメインで、とにかく事故を起こさないというのがゴールなお仕事です。なかなか返信をくれない講演者 (有名な研究者とか会社のマネージャー) たちに臆することなくリマインダーメールを投げまくるといったスキルが身に着きました。

研究の方面では、常識的知識を使った気の利いた対話システムの開発に取り組んでいます。今学期は今までの機械学習色が強めなものから少し方向転換をして、データベースを賢く探索しながら良い推論パスを見つけるという方法を試しています。最近の主流からは外れるので不安はありますが、同じトピックに随分長い時間をかけてきたのでそろそろ出口を見つけたいところです…。その他にも、私が昨年学会で発表したトピックである多言語単語ベクトルに関連する研究に取り組んでいる修士の学生をメンタリングしています。手法だけでなく、評価方法・データはどうあるべきかという話に踏み込んでいて、なかなか私自身にも学びが多いです。(と同時に、評価が適当な既存研究が結構数存在することに辟易したりもしています。)

5月からはマイクロソフトリサーチでインターンシップをしていて、Todo 管理アプリ (Microsoft To Do) に関係する言語処理プロジェクトに取り組んでいます。所属はシアトル・レッドモンドの研究所ですが、今年はリモート形式で、残念ながらインターンの楽しさ自体は半減という感じです。一方で、メンター+コラボレーター4人から毎日たくさんのフィードバックやアドバイスをもらえるのは張り合いがあって楽しいです。ここまで予備的な実験は珍しく想定通りに進んでいるので、この調子で色々なアイデアを試してみたいと思います。

3 NLP コロキウムはじめました

3月に自然言語処理・計算言語学に関する研究発表をオンラインで行う『NLP コロキウム』⁴というイベントを立ち上げました。⁵名前からわかるように、私が TA をした LTI Colloquium をマネして、研究

²COVID-19 ワクチン接種後にベル麻痺に罹患する確率が上がるという報告があるそう。有意な相関はないという報告も出てみたいなので、実際の真偽は不明

³<https://lti.cs.cmu.edu/lti-colloquium>

⁴<https://nlp-colloquium-jp.github.io/>

⁵京大学部の同期であり、最も信頼・尊敬できる友人の一人である横井くん (現東北大助教) と共同で企画しました。6月から東工大の博士課程学生で非常にアクティブに活動している丹羽さんにも参画してもらっています。

NLP コロキウム 

NLPコロキウムは自然言語処理 (NLP) および計算言語学 (CL) に関する最先端の研究・開発に取り組んでいる方をお招きするトークイベントです。

参加用URLなどNLPコロキウムに関するお知らせを受け取りたい方は[こちらのフォーム](#)からメーリングリストへのご登録をお願いします。

主旨

人工知能技術が著しい発展を遂げ、中でもコミュニケーションと知識表現のための主要メディアである言語の処理に世界中の注目が集まっています。多くの方がNLP/CLの研究・開発に従事するようになり、日本も例外ではありません。しかしながら、日本発の研究や日本語を対象とした研究の国際的なプレゼンスは研究者の層の厚さや日本語話者の数を考えるとかなり少数だと言わざるを得ません。

このトークシリーズを通じて 真剣にNLP/CLに取り組んでいる (あるいは取り組もうとしている) 研究者・技術者が所属や地理的な距離を超えて交流できる場をつくりたいと考えています。またこれによって、新しいコラボレーションやキャリアアップの機会が生まれること、最先端の研究トレンドに関する知見が共有

発表だけでなく発表者との個別ミーティングの機会も提供しているほか、許可をもらえたものに関しては発表の録画を YouTube で公開しています。講演者の招待、調整、動画の録画編集などなど、TA で培った経験が大いに役立っています。現在ほぼ隔週のペースで開催していて、毎回だいたい 60 人から 80 人くらいの方に参加していただいています。奨学生で NLP を研究している安永くんと笠井くんにも研究発表をしていただきました。感謝です！また、企画段階でもお二人を始めとした色々な人にアドバイスをもらいました。感謝感謝です。そしてイベントを盛り上げてくださる参加者の方々にも感謝しかありません。

オンライン開催というフォーマットには空間的な制約に縛られず誰でも参加できるという利点があります。このイベントを通して、日本のコミュニティ内の相互認知を高めて、日本全体の研究の質や世界的な存在感 (そして私が将来日本に帰ったときの日本の居心地…) の向上に寄与できたらと思います。今こうしてアメリカに留学している身として、少なくとも日本国内外を橋渡しして、海外に出るときの取っ掛かりを見つけやすくしたり、留学生が日本に一時帰国したときの楽しみを増やしたりするお手伝いをすることを目標にしています。

現状ではまだまだオーディエンス側からのアクションが少ないと感じていて、相互交流という面を日本の文化に合わせて今後改善していきたいと思っています。また、講演者をリストアップする段階で、同じような属性を持ったよく他所で講演している人ばかりを上げてしまうことに気づきました。このマタイ効果あるいは無意識のバイアスみたいなモノに抗うことも一つのテーマにしています。…とはいえ聴衆の期待にマッチしないと講演者共々不幸せになってしまうので、バランスが難しいです。

次の報告書を書く頃には一億人くらいの参加者を集めるイベントに成長させたいと思います。

4 おわりに

次の年度で、留学してから 5 年目に突入します。初めの 2 年は修士だったので、博士課程 3 年目と自称するのも許される風潮もありますが、指導教員の懐に甘えすぎず、そろそろ腰を据えて卒業までのパスを考えたいと思います。