

船井情報科学振興財団 第二回留学報告書

最初の学期を終えて

大谷直樹

2017年12月

2017年秋よりカーネギーメロン大学 (Carnegie Mellon University; CMU) の修士課程 (MS Program in Language Technologies; MLT) に在学している大谷直樹です。最初の学期は苦労も多かったですが、とても充実していました。

1 MS Program in Language Technologies (MLT)

まず私が在籍している修士課程について簡単に紹介します。奨学生の荒木さん¹、川上さん²(現在Oxford)、そして林くん³の報告書でも紹介されていますが、LTIは自然言語処理を専門に扱う研究所です。

私が所属する MLT は研究志向の修士課程で、他の修士プログラムの学生よりも少ない週5コマ (1コマ=80分) 程度を各学期履修することを求められます。すべての学生が何らかの研究プロジェクトに取り組んでいて、そのうちの多くの学生が Research Assistant (RA) や Teaching Assistant (TA) のポジションを得て学費免除と給料を得ます。



(a) School of Computer Scienceのメインの建物 (Gates Hillman Complex; GHC)。赤枠で囲割れている建物が LTI。



(b) GHC 内にある LTI の入り口。

¹http://www.funaifoundation.jp/scholarship/grantee_araki_jun.html

²http://www.funaifoundation.jp/scholarship/grantee_kawakami_kazuya.html

³http://www.funaifoundation.jp/scholarship/grantee_hayashi_hiroaki.html

MLT は毎年 30 人程度の学生を取るのですが、今年は 50 人近い学生が入学しました。例年は合格通知を多めに出しても学生が他の学校へ流れるそうですが、今年は MLT を選ぶ人が多かったようです。学生が増えても教員と RA のポジションの数は増えませんから、学期はじめの指導教員選びは熾烈でした⁴。

LTI は School of Computer Science の建物群の一角にあります。オフィスの配置はランダムで、なかなか先生の噂話も大声ではできません。PhD 学生は 3 人程度で一つの部屋をシェアしており、教員は一部屋が割り当てられています。修士の学生には固定席はなく、大部屋を全員で共有することになっています。ただし MLT は LTI の建物から少し離れたところに共有オフィスがあり、ちょっと不便です。最近では寒くなってきて極力外に出たくないで、私も大抵はメインの建物の共有スペースで過ごしています。

2 講義

今学期は 3 つの講義を履修しました。アメリカ大学院の講義は非常に大変という話を聞いていたので身構えていたのですが、やってみれば (たまたま?) 程よい負荷で、無事に学期を終えることができました。私は日本で修士課程まで修了していて、すでにある程度基礎がある状態で臨めたのが良かったかもしれません。

Algorithms for NLP⁵ NLP の代表的なアルゴリズムについて学ぶ必修講義です。内容は有名な教科書のひとつ『Speech and Language Processing』がベースになっています。

Machine Learning for Text Mining⁶ 情報推薦、リンク予測、文書分類などメジャーな課題と既存手法について学びました。

LTI Colloquium 毎週一人のゲストが招かれる招待講演です。様々なトピックの最新の話題を聞いて楽しめました。

3 研究

現在は言語資源が少ない言語に対して情報抽出や翻訳といった処理を行う技術を研究するプロジェクトに所属しています。誤解を恐れず大雑把に言うと、最先端の NLP 技術の多くは大量の正解データが存在することを前提としていて、それをいかに効率よくコンピュータに覚えさせるか、というところにあります。しかし、世界に何千種類と存在する言語のうち、整ったデータが存在するのはほんの数十種類の言語です。残りの大多数の言語にはディープラーニングのようなオシャレな方法は直接通用しません。それぞれの言語に対してデータを作れば済みそうですが、それには膨大なお金と時間がかかってしまい現実的ではありません。

いま、この課題に対して能動学習 (active learning) というアプローチで取り組もうとしています。これはコンピュータが自身の状態に応じて人間に能動的に正解データを要求するという枠組みです。ランダムに正解データを作るのではなく、本当に必要なデータを見極め、重点的にデータを作っていくことによって少ない資源構築コストで情報抽出や翻訳の精度を最大限に上げることを目指します。

私は以前からクラウドソーシングという主にインターネット上の不特定多数の人にアノテーションなどの作業を依頼する仕組みに取り組んできました。幸運にも今の研究トピック

⁴LTI では多くの場合入学後に指導教員が正式に決まります

はそれと多少関係していて、自分の経験が活かせるような気がしています。また、もともと興味があった多言語の処理を研究できるという点もワクワクしています。

こちらでは Yiming Yang 先生の指導を受けています。先生はテキストマイニングや情報検索、機械学習で著名で、ミーティングなどでも特に統計モデルに対する洞察の深さを感じます (個別ミーティングでは毎回モデリングのもっともらしさについてみっちり詰められます)。

同じグループの学生はほとんどが中国人です。グループセミナーは英語で行いますが、学生同士の日常会話は主に中国語です。私はいま一人の PhD 学生と一緒に研究をしていて、ミーティングではたまに英語交じりの中国語を使っています。おかげで私の中国語も少し上達した気がします。

4 京都大学での修士課程との比較

学費 CMU の学費は京都大学のおよそ 10 倍です (高い...).

講義 学費の違いを反映してか、今学期に履修した講義は京都大学ではなかなか見たことがないようなクオリティでした。特に Algorithms for NLP は素晴らしい講義でした。指定された教科書は一度読んだことがあったのですが、講義のおかげで理解が深まりました。また宿題が非常にタメになりました。Java を使って代表的なアルゴリズムを実装するのですが、アルゴリズムの理解が問われるだけでなく、効率を考慮したプログラムを書かないと合格点に達しないなど、とてもよく考えられた課題になっていました。また、TA が優秀でした。TA が行う週一回の講義では、学生とは思えない知識量と説明のうまさに驚きました。ただ、タメになるといつつも、講義の宿題や試験は実際かなりの時間を要するので、提出締切や試験前になると自分の研究の進捗に影響が出てきます。この点、京都大学の修士課程にいたときのほうが集中して研究に取り組むことができました。

学生の積極さ 講義では学生が積極的に質問をします。京都大学の講義では先生が質問しても講義室が静まり返るのが普通でしたが、こちらはしばしば早押しクイズ状態になります。明示的に問いを出されなくても、学生の方から質問をどんどん飛ばします。おかげで講義が予定より遅れることもよく起こりますが、それが普通のようなです。ミーティングでも学生は積極的に発言します。発言がないと何もしていないのと同じ、という認識すら感じます。

研究指導 全体的に京都大学のほうが教員一人あたりの学生が少なく、丁寧な研究指導をしているように感じます。こちらは教員一人で十数人の面倒を見ており、ミーティングの頻度も様々です。ただ、学生の研究能力が平均的に高いので、先生の助言が少なくても学生同士で議論してそこそこのレベルまで到達している感じもします。

研究プロジェクトのプレッシャー 多くの MLT 学生はプロジェクトで雇われています。しかも、来年に控える PhD のアプリケーションで勝ち残るためには成果を出して顔売らなければいけません。そうするとプロジェクトのタスクの優先度は必然的に高くなります。タスクは研究にならないエンジニアリングであることもあります。講義の課題をこなしつつプロジェクトの仕事もするのはなかなか容易ではありません。

5 アメリカで自動車免許を取る

最後にこちらでの生活について少しだけ紹介します。アメリカの美味しい地ビールについて書こうかと思いましたが、それはまたの機会にして、今回はアメリカで自動車免許を取る方法について書きます。私が京都に住んでいたときは、どこに行くのも大体自転車で移動していました。帰省するたびに早く自動車免許を取れと言われるのを頑なに拒み続けて6年、ついに一念発起しアメリカで免許を取りました。

アメリカの自動車免許試験は筆記試験と実技試験に分かれています。筆記試験に合格すると仮免許 (learner's permit) がもらえ、助手席に免許所持者が乗っている場合に限って自動車を運転できるようになります。その状態で自分で運転を練習したあと、自分で車を用意して試験場に持っていき、その車で実技試験を受けます。さすがアメリカ、無駄がないですね。全体を通して必要な料金は25ドル (+健康診断代金) でした。日本で取得するよりもずっと安く済むので、これから長期留学される方で免許を持っていない場合はこちらで取ったほうがいいかもしれません (車線の逆に適応する必要もなくなります)。

9月上旬	近所の病院で健康診断を受ける (体重・身長・脚や目に異常がないかなど)
10月13日頃	筆記試験の勉強を始める
10月21日	筆記試験に合格
10月22日	初めてハンドルを握った直後、路上に出てスーパーに行く (恐怖)
11月25日	実技試験合格

運転はすべてルームメイト (以下、教官) に毎晩指導してもらいました。STOPサイン無視、信号無視、ウィンカーを出した方向とは逆に曲がるなどやってはいけないことはだいたいやりましたが、教官の車を傷つけることなく、身体も無事のまま運転免許証を手にすることができました。ほぼ毎晩練習に付き合ってくれた教官には頭が上がりません。

筆記試験と実技試験の内容は州ごとに異なるようです。Pennsylvania は親切にも筆記試験対策用のアプリを公開しています⁷。実際の試験と同じ問題が出るというこの上なく実戦的な仕様になっているので、試験場で「あ！この問題、アプリでやったやつだ！」を実体験できます。実技試験は路上駐車が多い Pennsylvania の実情を反映して、縦列駐車が最初の試験項目に入っています。縁石にタイヤを当てたり、決められた線をはみ出た場合路上に出ることなく試験終了になります。

私はどちらも無事に通過することができ、つい先日免許証を受け取りました。まだ自分の車を買う予定はありませんが、これで晴れてパスポートを持つことなくビールを飲みに行くことができるようになりました。

6 おわりに

期待に胸を膨らませて Pittsburgh に来てからはや4か月、噂に違わぬハイクオリティの講義と期待を裏切らない優秀な学生に囲まれて張り合いのある日々を送ることができました。講義と研究のバランスや英語でのコミュニケーションがうまくいかず、思い通りのパフォーマンスが出せない悔しさを味わうことも多くありましたが、最近は鈍感になってきて、多少失敗してもゴールに近づいていればあまり気にならなくなってきました。来学期の目標は今学期以上に研究に時間と頭を使って、論文になるような成果を上げることです。

⁷PA Driver's Practice Test