

船井情報科学振興財団第六回留学報告書

大谷直樹

2019 年 12 月

2017 年秋よりカーネギーメロン大学 (Carnegie Mellon University; CMU) の Language Technologies Institute (LTI) に在学中の大谷直樹です。博士課程に編入して最初の学期を終えました。

1 近況報告

博士課程に入りついに LTI の建物内に自分のデスクをゲットしました。これまで 2 年間のノマド生活¹より遥かに快適な毎日を送っています。オフィスは別の研究グループに所属している LTI の博士学生と共有しています。どちらも非常に優秀かつワーカークホリック勤勉なので刺激になります。

私生活ではミートスライサーを購入して薄切り肉を食べまくったり、ピッツバーグのサッカーチームを応援したりしています (下図)。

2 研究

これまでの 2 年間で必要単位を取り終わったため、今学期は講義は受講せずに研究に集中しました。タイムマネジメント能力の向上のおかげか、単純に宿題がゼロになったのおかげか、学期を通して一晩も徹夜をしないという快挙を達成しました。

今年の春から、博士論文を見据えた研究テーマとして文書からの「常識」の獲得と、その言語処理タスクへの応用に取り組んでいます。文書に限らず口頭の会話でもそうですが、人間のコミュニケーション

¹修士課程在籍中も一応共有オフィスが使えましたが、アクセスが不便だったのであまり使っていませんでした。



(a) オフィス (ごちゃごちゃしている机はオフィスメイトのもの)。窓の外は歩道なので、たまに歩行者と目が合う。



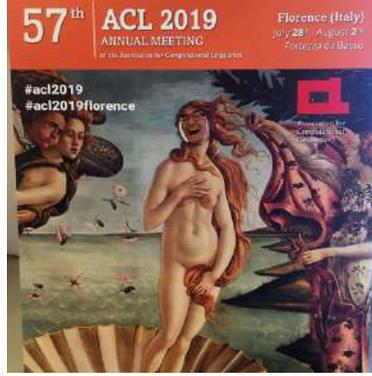
(b) アメリカのスーパーでは薄い肉が手に入りにくいので、塊を買って薄切りする。



(c) 試合後の風景。固まっているのが選手で、真ん中の観客席にいると思われるのがオーナー。ちなみに試合後はフィールドに入って選手にサインもらったりできる。小さいクラブなので...



(a) 学会の様子。3000人近くの参加者がいたそう。



(b) フィレンツェでビーナスとして誕生した私。産声を上げています。



(c) 元気そうに見えるが前の晩に飲みすぎたせいで水が手放せない。

はしばしば不完全です。例えば料理の指示を想像してみてください。卵を溶くためには、実際、卵を割る、中身を容器に入れる、殻は捨てる、といういくつかのステップを踏む必要があります。ところが、大抵の人は「卵を溶いてください」と指示されるだけで必要な手順を想像できると思います。大事な情報が欠落しているのにコミュニケーションが成立するのは、誰もが知っている事柄—常識—が暗黙的に使われていると考えられます。これが高度な言語理解をする上では非常に重要で、コンピューターにとっては厄介な問題です。また、この問題の難しさは、常識は常識であるがゆえに文書などの中に現れにくく、自動収集しづらいという特徴にもあります。私のプロジェクトでは知識の自動獲得と選別、そしてその一般化と応用のための方法を構築することを目的にしています。今学期は問題の定義とデータ作成だけで終わってしまいましたが、来学期以降は実際の知識の獲得と応用に取り組むつもりです。

夏は今年の春まで取り組んでいた内容について学会発表をしてきました。場所はイタリアのフィレンツェです(上図)。いつも通り楽しく充実した時間を過ごしました。今回の学会で見た発表のトレンドを私の好みを多分に加味して雑に分類すると、次の3つになるかなと思います。1つ目は、超大量のデータを使って学習した単語列のパターン²が様々な応用技術(テキスト分類、質問応答、翻訳)の精度が上がるという研究。直感的には、言語に関するタスクをこなすためには、やはり言語の単語や文法を知っておいたほうがいい、という感じだと思います。最近 Google が検索エンジンの中に関連する成果を導入したことが話題になりました。2つ目のトレンドは、倫理的な問題に対する考察です。言語処理技術はすでに人間に大きなインパクトを与え得る局面—テストの採点や履歴書の自動選別—で実用化に向かいつつありますが、データから取得したパターンを使う技術は、例えば医者やエンジニアの選考で男性を優遇するなど、社会(= データの出处)に存在する偏見を増長させてしまう可能性があります。言語処理コミュニティでは、こうした問題を緩和する方法が近年盛んに研究されています。最後のトレンドが私のお気に入り、技術の評価、の評価です。大多数の言語処理研究は、新しい技術を提案して特定のベンチマークデータセットで評価するというお決まりの流れで行われます。ベンチマーク上の性能によって様々な仮定の正当性を議論したりするわけですが、最近、よく知られたデータセットの中に誤りや予期しない“ヒント”が含まれたという報告が何件もあり、ベンチマーク自体を再評価するべきだという意見が強くなっています。品質の低いベンチマークが提案され無批判に使われてきてしまった背景には、ここ数年の研究スピードの加速や必要データサイズの増大があると思います。ベンチマーク作成は評価対象に関する深い洞察や丁寧な設計が必要な、本来かなり骨の折れる仕事です。こういう課題に注目が集まるようになったのはいい傾向だと思います。

今学期は本題の研究プロジェクト以外の仕事として、短期留学に来た学部生と LTI の修士学生のメンタリング、という「お兄さん活動」をしました。私が取り組んでいる研究プロジェクトを手伝ってもらい

²言語モデルと言います。

つつ研究の経験を積んでもらう (実質的には後者のウェイトが高め) という取り組みです。一番難しかったのが長期的な見通しを持ちながら適度な粒度・難易度・面白さのサブタスクに学生を誘導することで、かなり良い経験になりました。間違いなく今後の自分の取り組み方にも役に立つと思います。

CMU では学生向けの様々なワークショップが毎週のように開かれるのですが、ちょうど “Mentoring Undergraduates in Research Workshop” というワークショップがあったので参加してみました。新しい学びは沢山ありましたが、中でも印象的だったのが、実力と時間的制約に合わせたゴールを早い段階に学生と一緒に設定する、ということでした。指導される学生がゴールを意識できていないと、作業進度が落ちたり、作業の遅れに対して必要以上の心理的なプレッシャーを感じてしまうことになりがちです³。このアドバイスを参考に、学生とミーティングするときにはしばらく先までの予定を踏まえて、最低限達成したいことなどを設定しました。学期末はやはり週末を献上しなければいけないほどに忙しくなってしまったのは反省ですが、全体的に楽しく研究できました (と、メンタリングした学生たちも同じように思っていることを祈って...)

3 おわりに

自分の研究を進める以外に学生のメンタリングや共同研究者との調整などの仕事が増え、博士課程に進んだ実感を感じています。来学期も健康に気をつけて楽しく頑張ります。

³学部生であれ修士学生であれ、研究研究を積む・成果を出すということがキャリアに響くため皆比較的モチベーションが高く、こちらでは後者の問題のほうが大きかったりします。