

船井情報科学振興財団第七回留学報告書

大谷直樹

2020年6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。まもなく留学生活も3年になろうとしています。

1 近況報告

今頃インターンのためカリフォルニアで最高の夏を過ごしているはずでした。残念ながらリモートワークで行うことになってしまい、相変わらずにわか雨の多いピッツバーグで毎日過ごしています。カリフォルニアに行けなかったのは今世紀最大の悲しみですが、それ以外は問題なく、むしろ毎日運動しているおかげでここ数年のうち最も健康な生活を送っている気がします。

意識的に運動を始めた理由はカリフォルニアのビーチに似合う立派な肉体を手に入れるためでした。昨年末インターン先が決まってすぐジョギングと筋トレを始めたのですが、間もなく海のないピッツバーグでの幽閉が決まり、完全に目標を失ってしまいました。3月頃からはこの肉体の使い道を求めてテニスを始めました。途中、スライスボールを変な姿勢で取ったときに腰の爆弾が爆発して一週間ほど歩行困難になったりもしつつ、今に至るまでほぼ毎週1-2時間ほど友達と打ち合っています。6月に入ってからはたまにゴルフにも行くようになりました。健康的すぎて死にそうです。

それ以外のイベントとしては、同じアパートに住む日本人と一緒におせちを作ったり、初めて銃を撃ったり、ピッツバーグでの暴動でパトカーが炎上した跡を見学したりしました(下図)。充実した毎日です。



(a) アメリカで過ごす正月。手作りおせち料理が非常においしい。



(b) 生まれて初めて銃を撃っています(実弾)。なお弾は当たりません。



(c) デモを発端とする暴動でパトカーが燃えた跡。道路が焦げてます。

2 研究

3月の半ばから学校の建物に入ることが禁止になり、以来ずっとアパートに自室で研究・学習をしています。生物や化学系の研究と異なり、パソコンとインターネットがあれば私の研究は場所を選ばずできるので、幸か不幸か研究が止まるということはありません。ただ、研究ができていても進度のほうはイマイチで、特に行き詰まったときの気分転換が自宅にいるとうまくできず、苦勞しています。

研究としては、これまでに引き続き、人間が持つ「常識」が言語コミュニケーションの中で担う役割の分析、常識の獲得とその応用に取り組んでいます。人間が持つ知識を組み込んで言語処理技術をより高度にするというのは非常にホットな研究課題のひとつです。というのも、人間の情報伝達では、言葉として明示的に現れない背景知識や共通理解がしばしば重要な役割を果たすからです。例えば、「トロフィーはスーツケースに入らなかった。大きすぎたからだ。」¹という文で、大きすぎたのは何でしょうか？人間にとって簡単な問題です。「物体を容器に入れるためには、その物体が容器より小さくなくてはならない」という暗黙的な事柄を知っているから、私達は大きかったのはトロフィーだと理解ができるのです。このように、言葉として明示的に現れない情報が言語理解のキーになっているケースでは、必要な知識を持たないデータ駆動の機械学習技術はうまく正解に辿り着くことができません。

毎年、Knowledge や Common-sense に関するいくつかの論文が言語処理技術系の学会で発表されます。多くは、機械学習モデルを構築する際に機械が扱いやすい形式に構造化された知識リソース—例えば、(バナナ, IsA, 果物) や (フォーク, UsedFor, 食事) のような事柄を記録したデータベース—を組み込むという方法を採用しています。知識リソースを入れない場合に比べて言語理解タスクでの性能が上がるというのですが、問題の本質からまだまだ遠いところにあります。結局現段階の成果は、知識にもとづいて言語を深く理解しているというより、あるパターンに対して反射的に特定の値を返すというパターン認識の域を出ていないと思います。

常識にもとづく言語理解は何十年も前から広く研究されていて、初期は論理学にもとづく計算アプローチが主流でした。今学期いくつか文献を読んでみたのですが、最近のデータ駆動な研究とは作法が違って読んでいて面白かったです。緻密な現象の分析、論理学モリモリな手法、オモチャみたいな小規模実験(あるいは実験全く無し)という流れで書かれていて、(ほぼ)統計的モデリング一辺倒で個別の事象分析より大規模実験の結果が重視される最近の論文とは見た目からしてかなり異なります。私にとっては以前の論文のほうが個別分析から手法をデザインする部分に含蓄があったのですが、その手法は単純化されたごく限られた問題設定でしか使えなかったり、開発に非常に手間がかかったりして実用に欠ける印象を受けました。

私は古典的な研究からエッセンスを汲み取って今風な問題・手法を提案することを狙っているのですが、なかなかうまく行きません。今年の後半では何らかの成果に繋がってくれることを祈っています。

研究以外では、自分の理論的知識を増やすべく言語学関係の講義を2つ聴講しました。ひとつ目の講義は主語や目的語といった単語間の関係を研究する統語論 (syntax) に関する講義で、もうひとつのセミナー方式の講義では文脈や話者と言語表現との関係を研究する語用論 (pragmatics) を学びました。後者は常識が意思伝達の中で果たす役割なども対象としており、今後の自分の研究に役立ちそうでした。

3 おわりに

しばらく外で自由に活動できない日々が続きそうですが、もともと引きこもるのが好きな自分にはあまり悪影響はなさそうです。週末の楽しみだったサッカー観戦ができないのが辛かったので、来シーズンは無事にリーグが開催されることを願っています。

¹Winograd Schema Challenge という著名な問題から引用。