

船井情報科学振興財団 留学報告書

林佑明

2020年7月

先の報告書から大変長い期間を空けてしまいました、カーネギーメロン大学言語科学研究所博士課程 (CMU, LTI) に所属する林です。

1 近況・研究

ピッツバーグには修士の頃からいるため、授業履修などはすでに終え、ここ何年かは研究に集中していました。昨年行った研究は、機械学習による文章生成のモデルに「知識」を組み込むことでした。ここでいう知識とは、例えばウィキペディアに載っているような「バラク・オバマは大統領である」、「バラク・オバマの妻はミシェル・オバマである」といった、物事・人物間の事実関係を表した辞書のようなものです。現行の自動言語生成モデルには、こういった決まった事柄を100%間違えずに生成する能力はありません。そのため、生成した文章が人が書いたように見えたとしても、事実関係がめちゃくちゃであることが多いです。それを克服するために通常の生成方法とは異なる、事実関係に基づいた言語生成の仕方をモデルに与え、機械学習を通して最適な生成手順を学ばせようという趣旨の研究を行いました。これにより、例えばモデルが「バラク・オバマは...」に続く文章を生成する際、事実と異なることを生成し辛くなります。

昨今の政治戦略（特に米国において）ではソーシャルメディアがよく使われ、どこまで本当でどれが根拠のない情報かを取捨選択する必要があります。その際に、文章が既存の事実関係と矛盾していないか、あるいは事実関係と矛盾しない文章を生成することで正確にニュースの内容を伝えたりできるか、といった点で研究が活用されれば嬉しいです。

最近では文章要約生成の研究を行っています。大学入試で出題されたことがある大学生は多いと思いますが、文章要約というのは長い文章から重要な点を抽出したり自分なりの言葉でまとめることで、短く内容を伝えることを指します。優秀な要約モデルがあれば、例えばこの報告書を一文に要約してパッと読み終えることで読み手の皆さんの貴重な時間を短縮することができます。

要約モデルにおいて大事なのは、「重要な点」の選別基準です。例えばレストランのレビューを読んでいる場合、人によって金額、味、雰囲気、など違った基準に重きを置きます。そうした基準ごとに要約を生成させる研究を行っています。言語処理のコミュニティでは「重要な点」を普遍的に重要な情報として扱い、要約モデルを構築する研究が今でも盛んに行われていますが、基準・目線を変えた状況下でどのようによい情報の要約を得るかの研究はあまり盛んではありません。したがってデータの量も少なく、大

規模なモデルを学習するのはあまり得策ではありません。私の直近の研究ではこれを解消すべく、大規模なデータを用意して、その上で要約モデルを作っています。

2 コロナウイルスによる自粛の影響

ここまで長引くとは予想もしていませんでしたが、コロナウイルスの拡大に伴い3月から自宅で研究をしています。実験系の研究と違いパソコンとWiFiでなんとかなるため、むしろそのおかげで自宅でどれだけでも長く研究できます。そのため、「休む・働く」をしっかりと区別することが難しくなり、ダラダラしてしまったりすることがありました。今でもバランスを意識して毎日を過ごしていますが、時々崩れてしまうことがあり大変です。

外出を基本的にしなくなったので、週に何度かの散歩がなければ全く体を動かさないことになります。大学と家の往復程度の運動もしなくなってしまう危機感を感じたため、最低限の筋力トレーニンググッズを買い、思い立ったら体を動かすようにしています。また、マスクをして人とのすれ違いを避けながら散歩をするようにしています。

先学期は3月以降授業もオンラインになりました。そんな中私がTAで関わっている授業で一コマ講義をする機会をもらい、初めてリモート授業なるものをして、いくつか気づきを得ました。まず、他人を目の前にすることがないことで、アガらずに済むということです。私は、どんなに練習しても人前で発表する能力は発展途上なので、画面に向かって喋り続けるのはとても楽に感じました。次に、原稿を用意しても一切見ながら授業をすることができなかつたということです。何らかのメモを手元においておけば、一呼吸おく間に目を通して話す内容を確認できると思います。私はそれをPCのメモに書いてしまい、それが発表中のスライドで隠れてしまったので、結局全く見れませんでした。重要な発表を今後する時には、モニターをいくつも用意してカンニングペーパーを物理的に手元に用意しておくべきだと学びました。

あとは買い物のある方に変化がありました。基本的に外出もしないので、Whole Foods (Amazon) や Instacart を使ってインターネットで生鮮食料品を注文します。届いた通知が来たらすぐに受け取りに行かないと、要冷蔵のものがだめになってしまうので大変です。このシステムで一番だめな点は、品物(とくに野菜・果物)を直接見て選べないので、買ってきたものの一部が既に傷んでいた場合は受け入れるほかないところです。それでも時々自ら買い物をしに行きますが、お互いマスクをしていてもなるべく離れたり、どこに手を触れるか気を遣ったりと、神経質になりながら外出しています。

3 おわりに

コロナウイルスに加えて、オンライン授業に起因するビザの取り消し問題も出てきており(7/6/2020)、ますますストレスの種が増える一方ですが、そんなときは重りを持ち上げて、筋肉とともにストレスをエネルギーに変換していきたいと思います。