

船井情報科学振興財団 留学報告書

2017年6月

荒木 淳

1. はじめに

私は、2012年8月よりカーネギーメロン大学コンピュータサイエンス学部言語技術研究所 (The Language Technologies Institute of the School of Computer Science at Carnegie Mellon University) の博士課程に在籍しています。今学期 (2017年春学期) は博士課程の最終段階にあるのですが、今までの中で研究課題に最も四苦八苦した時期でした。今回の報告書では、研究とその他の活動として論文の査読と大学入試の面接官ボランティアについて触れます。

2. 研究

私の博士研究では、テキストからのイベント構造の自動抽出という課題に取り組んでいます。もう少し具体的に言えば、テキストからイベント表現 (例えば「買った」や「地震」など) とその関係 (例えばある文中の「買った」と「購入」という表現の関係) を自動的に抽出するコンピュータモデルを考案する、という研究課題です。この課題の難しさの一つは、人手によるイベント表現の注釈付けはコストがかかるため、既存のデータセットのサイズが小さく、自然言語処理の分野で通常よく使われる機械学習の手法 (教師あり学習) では高い抽出率を達成することが難しいという点です。昨年11月の博士論文プロポーザルを終えた後からその課題に本格的に取り組み、試行錯誤の日々が続きました。様々な手法を模索したのですが、最終的にはそれらの手法で使った要素技術や言語資源を組み合わせ、そこから一つ飛躍的な思い付きがあり、解決の糸口を見つけることができました。

3. 論文の査読

昨年の秋頃から論文の査読 (レビュー) をする機会が増え、これまでに2つの国際会議で計9本の論文を査読しました。論文を様々な観点から評価することは批判的考察と建設的な意見出しという研究スキルを培う良いトレーニングになりますが、同時にその論文で主張されている最新のアイデアやその書き方からも得られるものがあります。この後者の意味での収穫は、研究分野によるかも知れませんが、基本的には一流の国際会議や論文誌に投稿された良質な論文を査読した方がより得られるものが多いと思います。一流の国際会議や論文誌に投稿された論文の質も様々ですが、他の国際会議や論文誌に比べて良質な論文に巡り会える可能性が比較的高いです。このためには、(1)自分の研究分野で一流の国際会議や論文誌に論文を通し、(2)学会などの場で研究者とのコネクションを増やし、(3)(知り合いの研究者から査読依頼を受け)一流の国際会議や論文誌での査読をしていく、という積み重ねが大事だと思います。この3ステップの中で(1)の継続性が最も大事で、私自身まだまだなので、今後もこの点を向上できるように努力していこうと思っています。

4. スタンフォード大学の大学入試の面接官ボランティア (続編)

前々回の報告書に少し書きましたが、以前スタンフォード大学から大学入試 (undergraduate admission) の面接官ボランティア¹に卒業生として協力してくれないか、という依頼を受け、

¹ <http://admission.stanford.edu/alumni/oval/interviews.html>

受諾しました。面接官トレーニングのためのワークショップに参加した後、今年の2月に初めて面接官を担当しました。面接のガイドラインや詳細な内容などは口外できませんが、スタンフォード大学が入学してくる大学1年生に対してどのような資質を求めているのかという点は興味深かったです。そもそも人物評価というのはバイアスがかかりがちで難しいもので、それを面接という限られた時間の中で英語で行うのは私にとってはさらに難しいことです。また、卒業生による面接の評価が合否決定に対してどの程度考慮されるのか分かりませんが、大学側の評価とあまりに違っているとその卒業生の信用にも関わり得るので、面接する側にも相応のプレッシャーがかかります。その難しさとプレッシャーの中で面接官を担当したことで私自身にも学びがありました。