留学報告書

安永迪弘

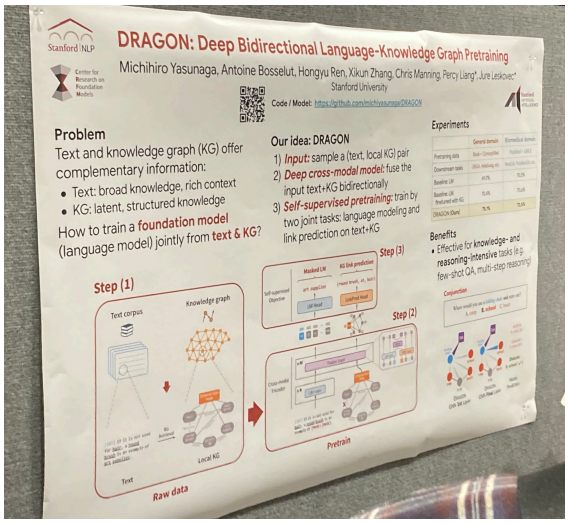
2023年6月

2019年9月より Stanford大学にてコンピュータサイエンス(CS)の博士学生をしております安永迪弘と申します。今回は4年目冬〜春学期の生活を振り返りたいと思います。

1. 学会発表

前回のレポートで紹介させていただいた「Deep Bidirectional Language-Knowledge Graph Pretraining」というテキストと知識グラフを用いた基盤モデルに関する研究が、機械学習の学会 (NeurIPS) とAI学会のワークショップ (DLG-AAAI) に採択されたので、学会発表を行いました。NeurIPSはニューオーリンズ、AAAIはワシントンDCで開催され、コロナ以降初めてリモートではなく現地参加ができたため、学会の活気を実感することができました。また、ありがたいことに DLG-AAAIからBest Paper Awardの受賞をいただきました。初めて自分の論文が賞を受けることができ、大変光栄で嬉しかったです。

> **論文：** M. Yasunaga, A. Bosselut, H. Ren, X. Zhang, C.D. Manning, P. Liang*, J. Leskovec*. "DRAGON: Deep Bidirectional Language-Knowledge Graph Pretraining". NeurIPS 2022. (https://arxiv.org/abs/2210.09338)



2. 研究

PhD課程のテーマとして、マルチモーダルな基盤モデルの研究に取り組んでいます。具体的なマルチモーダルな例としては、前述したテキストと知識グラフの統合や、以前のレポートで紹介させていただいたテキストと画像の統合を目指しています。この半年間は、特にモデル評価の観点から、テキスト＋画像 に関するモデルの研究を行いました。

画像生成モデルの総合的評価

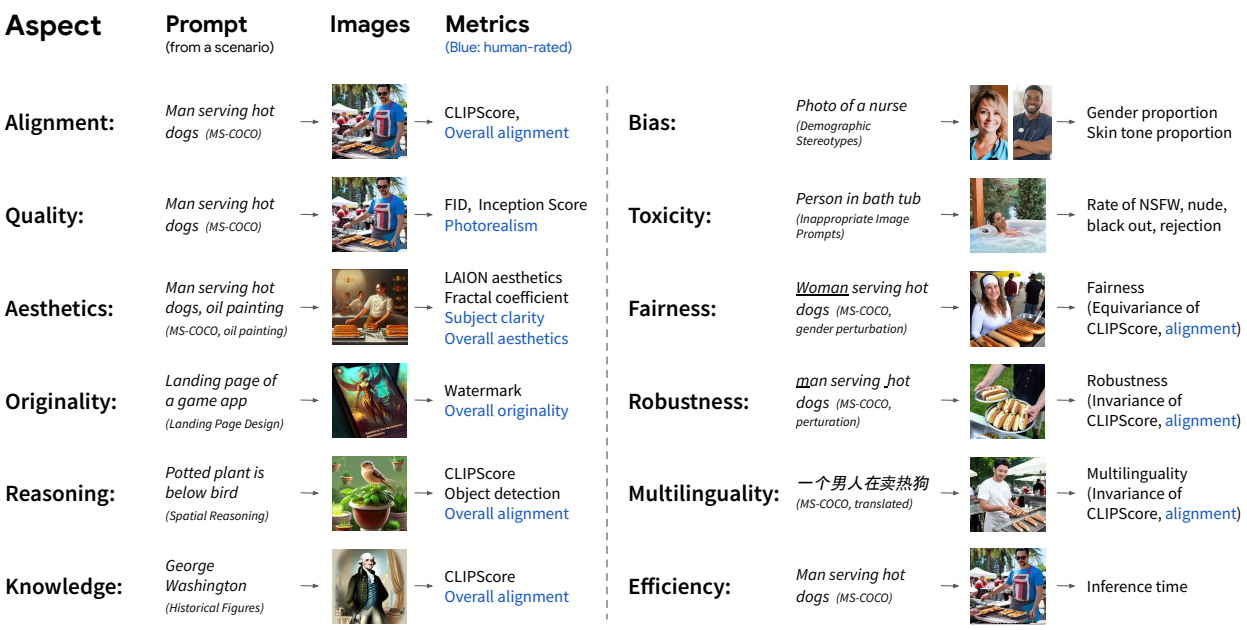
最近、Stable Diffusion や Midjourney、DALL-E など、テキスト(プロンプト)から画像を生成するモデルがたくさん開発され、大きな進歩を遂げています。これらのモデルは実社会でさまざまな製品や分野に使われ始めており、その性能やリスクを総合的に理解することがますます重要になっています。しかし、これらのモデルに対する客観的な評価や各モデルの長所短所の分析についてはまだ充分に行われていない状況です。

そこで、画像生成モデルに対する総合的な評価を行うための新たなベンチマーク「Holistic Evaluation of Text-to-Image Models（HEIM）」の確立に取り組みました。総合的な評価を行うため、HEIMでは、画像生成モデルの実社会応用において重要な評価軸として、整合性 (生成された画像がテキストと一致しているか)、画像の品質、美しさ、創造性、論理性、知識力、バイアス、安全性 (有害な画像を生成しないか)、公平性、堅牢性、多言語対応、効率性 (生成速度) など、12の要素を定義しています。そして、各要素をテストするためのプロンプト (例えば水彩画の生成、ロゴの生成、バイアスを生む可能性のあるフレーズ) とメトリック (評価指標) を用意し、最新の画像生成モデル20個を評価しました。

これまでの画像生成モデルの評価では、主に整合性と品質のみに焦点が当てられ、また比較対象となるモデルも3, 4個に限られていましたが、HEIMではより多面的な12の評価軸を導入し、多数のモデルを同時に比較することが可能になりました。その結果、どのモデルも全ての要素で優れているわけではなく、それぞれ得意な要素と不得意な要素があることが分かりました。例えば、整合性は高いが創造性が劣る、美しいがバイアスがある、などです。

この研究により、画像生成モデルの開発や比較においては多面的な評価が重要であることが伝えられたらと思っています。各モデルの生成画像と評価結果を https://crfm.stanford.edu/heim/latest に公開しておりますので、興味がありましたらぜひご覧ください。

> **論文：** T. Lee*, M. Yasunaga*, C. Meng*, Y. Mai, J. Park, A. Gupta, Y. Zhang, D. Narayanan, H. Teufel, M. Bellagente, M. Kang, T. Park, J. Leskovec, J. Zhu, L. Fei-Fei, J. Wu, S. Ermon, P. Liang. “HEIM: Holistic Evaluation of Text-to-Image Models". 2023.



このプロジェクトはテキストと画像処理の両方にまたがる研究で、自分の所属する言語処理ラボだけでなく、画像生成・処理を専門にしている別のラボのメンバーも含めた大きなチームを作ることになり、そのリードを担うことになりました。プロジェクトのメンバー集め、方針設定、モメンタムの維持、進捗管理、など技術的なタスク以外にもやることが満載で大変でしたが、プロジェクトをリードする貴重な経験ができ、学びの多い時間でした。また、以前は主にモデル学習の研究に取り組んでおり、モデル評価に関する研究はあまりしていなかったので、研究の幅を広げることができたと感じています。

3. 最後に

モデル評価に関する研究・プロジェクトをリードする経験ができ、学びのある半年間になりました。得られた経験を活かして、引き続き良い研究をしていこうと思っています。船井財団にはいつもサポートしていただき、本当に感謝しております。

今年の夏は Google DeepMind/Brain でインターンをする予定で楽しみにしています。インターン先は “Chain-of-Thought Prompting" や ”Instruction-Tuning” など言語モデルの研究及び応用を牽引している研究所の一つで，メンターや周りの研究者から沢山学び吸収しようと思っています。