

視覚と言葉を通じたロボットとの共同作業を目指して

日時：2021年2月17日（水）10:00～17:00

会場：遠隔配信

参加者数：99名

オーガナイザー：牛久祥孝（OSX / Ridge-i）

サブオーガナイザー：打田正樹（鈴鹿工業高等専門学校）

セミナー概要

最近の深層学習の発展は、コンピュータビジョン分野や音声信号処理分野、自然言語処理分野などにおいて画像認識、音声認識、機械翻訳や対話といった既存の課題に対するブレークスルーとなりました。しかし、深層学習によるもうひとつの重要な恩恵として、諸課題のために各分野で独自に発展してきた技術体系が深層学習という共通技術を取り込んだことで、異分野間での相互参入障壁が低下したことも特筆すべき変化です。結果として、画像や動画と自然言語を行き来するビジョン&ランゲージと呼ばれる分野も確立し、ロボットとの対話的な処理に視覚データを活用する研究など、人同士であるかのように人とロボットが共同作業できる未来に近づく取り組みが広がっています。本セミナーでは、こうした多様な分野で研究を進める講師の皆様から、最近の動向についてご紹介いただきました。

第1話 深層学習における判断根拠の視覚的説明と活用

中部大学 山下 隆義 先生

深層学習は画像認識分野において高い性能を達成し、認識する際の判断根拠の説明性について注目が集まっています。そして、判断根拠を視覚的説明として可視化する研究が進んでいます。視覚的説明は、人が深層学習の判断根拠を理解しやすいような形で可視化できるだけでなく、認識精度を向上させたり、人の知見を深層学習に導入したりできます。

本講演では、Attention Branch Networkによる判断根拠の視覚的説明について紹介し、人の知見を取り入れる仕組みや強化学習への応用も合わせて紹介しました。画像診断のように画像を見て判断のみを行うタスクから、実際に移動/運転したり糸を縫ったりと言った幅広いタスクにおいて視覚的説明の機構を活用されている事例を学ぶことができました。

第2話 音声認識の基礎と音声合成

東京大学 高道 慎之介 先生

音声は、喉・口を動かして生成される観点で物理的であり、言語・感情の伝達手段という観点で情報的です。故に音声分野において、信号処理と機械学習が古くから共存してきました。深層学習の発展は、分野間での相互参入障壁を低下させるのみならず、音声分野における物理・情報を享受した新たな枠組みを提供し続けています。

本講演では、音声の認識と合成を例にとり、最近の研究動向、異分野とのつながり、これから解くべき問題について触れて頂きました。人間の発声における解剖学的な導入からの音声認識や音声合成の解説や、音声と画像を含めたマルチモーダルな研究の事例もご紹介頂きました。

第3話 自然言語処理とビジョン&ランゲージへの派生

NTT 西田 京介 様

BERT に代表される、巨大なニューラルネットワークを大量のテキストで自己教師あり学習した汎用言語モデルの登場により、一部のベンチマークセットではAIが人間を凌駕するなど自然言語処理にパラダイムシフトが起きました。このアプローチは画像と言語の融合領域にも導入され、画像に対する質問応答などのタスクで大きな成果を挙げています。

本講演では、自然言語処理およびビジョン&ランゲージについて最近の研究動向と今後の展望について触れて頂きました。特に深層学習による自然言語についての導入を最近のTransformerまで丁寧に解説頂いた後に、画像と自然言語を含めたビジョン&ランゲージについて体系だっでご解説頂きました。

第4話 対話における画像・音声・言語情報の利用

理化学研究所 吉野 幸一郎 様

第3話でご紹介頂いたように、自然言語と画像情報を用いた Visual Question Answering などのビジョン&ランゲージから派生して、画像情報・視覚情報などをより広いコンテキストとして対話に用いようとする研究が行われるようになっていきます。この際、画像・音声・言語情報をどう効率的に利用するかが重要となります。また、実空間におけるロボット対話などでは観測される画像の情報をを用いることが必要不可欠です。

本講演では、こうした対話研究におけるマルチモーダル情報の利用について概観し、具体的な研究をいくつか詳説頂きました。マルチモーダル情報の活用方法の体系化から実際にそれらの事例としてのQA型対話や対話的画像編集といった事例紹介から対話制御の概要、そして実ロボットを用いた対話の研究まで幅広いテーマについてご紹介頂きました。

まとめ

今年度は第 125 回ロボット工学セミナーから、新型コロナウイルスの影響で完全オンラインでのセミナーを提供しております。もともと現地聴講と webex を用いた遠隔聴講を組み合わせた形式でセミナーをご提供してきているので、配信についての一定のノウハウはあったといえます。一方で、講師の先生方も遠隔で講演されること、現地で質問を受け付ける代替手段を検討する必要があったことなど、完全オンライン開催に伴ってより一層の検討と試行を繰り返しているところもあります。

今回は、私が担当した第 126 回と同様に sli.do で質問を受け付けました。匿名で質問できることや、他の聴講者の質問で自分も気になる質問には「いいね」を押すことでよりその質問に答えてほしいという意思表示が可能なことなど、オンラインならではの質問の形態ができる点はオンラインになったメリットの一つであると言えます。今回からは講師の先生方のご承諾も頂けたため、当日の聴講が叶わなかった方も含めてセミナー後の週末までの見逃し配信が提供されました。こちらも、オンライン配信をしているからこそプラスアルファの稼働を余り割くことなく、より多くの方々にセミナーをお届けできる方法だと思います。来年度のセミナーもしばらくは完全オンラインが続くと思われそうですし、またその後オフラインのセミナーが復活しても元来と同様にオンライン配信都のハイブリッドになると予想しています。オンラインならではのコンテンツをよりブラッシュアップすることと、ハイブリッドになる際は更に工夫することが求められているように感じました。

謝辞

それぞれの分野を代表する研究者の方々に、ご多用のところご快諾の上ご講演を頂いたことにまず感謝申し上げます。またこのセミナーにご参加頂いた参加者の皆様、質疑やアンケートに積極的にご参加頂いた皆様にお礼申し上げます。

また何より、事業計画委員会の委員の皆様には通常 1 人の委員が 1 回のセミナーをオーガナイズするところ、2 回のセミナーを企画・進行する機会を賜りありがとうございました。去年度に企画として自分自身では甲乙つけがたいものを 2 つ思いついてしまい、やはり需要の多い画像処理に特化したテーマ（第 126 回）に絞るべきかと思っておりましたが、両方やって良いというアドバイスを頂いた結果として、今回のセミナーをオーガナイズさせて頂く運びとなりました。サブオーガナイザーの打田先生にも当日の進行についてお力添えを頂きましてありがとうございます。また最後に、ロボット学会事務局の皆様にはセミナーの円滑な運営のために並々ならぬサポートを頂きました。感謝申し上げます。

2021 年 3 月 30 日

文責 牛久祥孝 (OSX/Ridge-i)