

TSUBAME 共同利用 平成28年度 学術利用 成果報告書

理解・制御可能なニューラルネット言語生成モデル
Interpretable and Controllable Neural Net Language Generation Models

中村 哲
Satoshi Nakamura

奈良先端科学技術大学院大学
Nara Institute of Science and Technology
<http://www.naist.jp/>

邦文抄録(300 字程度)

本研究では、ニューラルネットによる様々な言語生成モデルの学習において、どのような処理が行われているかを明らかにすることを目的とした実験を行った。さらに、言語生成におけるモデル化と探索アルゴリズムを開発した。

英文抄録(100 words 程度)

In this study, we did our experiments to discover the learning process of various kinds of language generation model using neural network. We developed a modeling method for language generation and discovering algorithm.

Keywords: 言語モデル, ニューラルネット, モデル化, 探索アルゴリズム, 言語生成

背景と目的

自然言語生成とは、何らかの入力を受け取り、その入力に基づき自然言語の文を生成する技術である。従来、言語生成を行う際に、その文を構成する要素(単語や句)を個別に生成し、これらを組み合わせて文を生成する要素合成手法が利用されてきた。しかしこの数年、従来の要素合成と全く異なる、深層学習とニューラルネット(NN)に基づく言語生成モデルが台頭し、多くのタスクにおいて、大幅に正確で流暢な文の生成に成功している。しかし、NN 言語生成モデルは平均的に滑らかな文の生成を実現する一方、予期せぬ深刻な誤りを起こすこともある。例えば、我々が現在開発中の科学論文用 NN 翻訳システムに「我々は直径を計った」と入力した際、学習データに十分カバーされていない「直径」を正解の「diameter」ではなく、「半径」を意味する「radius」に翻訳してしまい、重大な誤解を招く誤訳となってしまった。しかし、NN 言語生成モデルは入力を人間の解釈が困難な実数ベクトルとして表すため、内部でどのような処理を行っているか理解しにくく、予期せぬ誤りを起こした際に修正する手法が確立されていない。本研究は、まずこのようなニューラルネット言語生成モデルでどのような処理が行われているかを調査する目的にし、これを行うた

めに TSUBAME の大規模計算環境を利用して、様々な言語生成モデルの学習と比較を行う。また、この調査で得られた知見に基づき、ニューラルネットに基づく言語生成におけるモデル化と探索アルゴリズムを開発し、人間の知見を簡単に取り入れるニューラルネット言語生成の枠組みを目指す。こうすることにより、予期せぬ誤りを起こした場合に修正が可能となり、多くの応用においてより安心して使える言語生成システムへ繋がる可能性がある。

概要

自然言語の文を処理し、生成する技術は、人間の言語同士を翻訳する機械翻訳、人間と自然言語で対話を行う対話システム、画像から説明文を生成するマルチモーダル文生成など、様々な応用で用いられる汎用的な技術である。近年ニューラルネット(NN)に基づく言語処理モデルは、従来のモデルを大幅に上回る精度を実現し、注目を浴びている。しかし、これらのモデルには、内部でどのような情報が表現されているか、予期せぬ誤りを起こした場合、この誤りを人手で修正することが容易ではないという深刻な欠点もある。本研究では、様々な NN 言語生成モデルに対して、それぞれの特徴を調査

し、内部でどのような情報が表されているかを明らかにしてから、この調査に基づいて新たな制御可能なニューラルネットワーク言語生成の枠組みの確立を目指した。

結果および考察

実装上の問題により、TSUBAME 上で動作させることのできるプログラムの作成作業が完了しなかったため、言語生成モデルの特徴を解析することができなかった。この問題の原因として、GPUを利用したアプリケーションを構築することが困難であったことが挙げられる。

まとめ、今後の課題

本研究では、言語生成モデルの特徴を解析し、ニューラルネットの内部状態を明らかにすることを目的とした実験を行う予定であった。ところが、実装上の問題からプログラム作成作業が完了せず、目的を達成することができなかった。

今後は、このプログラム実装を完了させ、再び実験を行う。