

審査の結果の要旨

氏名 鈴木 雅大

この学位請求論文「深層学習と生成モデルによるマルチモーダル学習に関する研究」では、複数のモダリティをもつ情報に対して予測や推論等を行うマルチモーダル学習に関して、近年注目を集めている深層学習との統合を試みるものである。特に、深層学習と生成モデルを組み合わせた深層生成モデルのなかでも、変分オートエンコーダ (VAE) と呼ばれる手法に注目し、さまざまな新しい手法を提案するとともに、実験を通じて得られた知見を総合的にまとめている。

近年、深層学習が注目を集めている。画像認識や音声認識、言語処理などで顕著な性能の向上が報告されている。また、画像にタグがついたデータ、画像にキャプションがついたデータ、映像と音声など、複数のモダリティを持つデータについても深層学習による活用が進んでいる。一方で、深層生成モデルと呼ばれる生成モデルを深層にする技術も急速に進展している。VAE と GAN (Generative Adversarial Networks) がよく研究されているが、特に VAE は確率的な解釈がしっかりしており、基盤技術として重要である。マルチモーダルなデータに対して、深層生成モデルを適用した研究は、これまでにほとんど行われていない。

本論文では、マルチモーダル学習に関して、さまざまな問題設定における新しいモデルを提案し、その有効性を検証する。5章に述べる1つ目の研究では、JMVAE (Joint Multimodal Variational AutoEncoder) という手法を提案する。従来の手法では、双方向モダリティを変換する場合に、情報量の大きいモダリティを欠損させると共有表現が崩れてしまっていた。そこで、本研究では、単一のモダリティの場合と、モダリティを欠損させた場合の確率分布を近づけることによりこの問題を解決した。複数のデータセットによりその有効性を検証した。

6章に述べる2つ目の研究では、複数のモダリティから目標ラベルを予測する問題を扱っている。異なるモーダルデータはセットで手に入りやすいが、ラベルは手に入りにくいいため、半教師あり学習が有効である。本研究では、VAE

による半教師モデルをマルチモーダルに拡張した **Semi-supervised MVAE (Multimodal VAE)**と、それに複数のモダリティを統合する隠れ変数を追加した **Semi-supervised HMVA** を提案した。こうした手法では、モダリティの欠損が補完でき、単一のモダリティでも、また両方のモダリティが与えられた場合にも高い精度を実現することを示した。

7章に述べる3つ目の研究では、ゼロショット学習を扱う。ゼロショット学習は、一度も学習したことのない目標カテゴリのモダリティを、他の目標カテゴリでの学習結果と、異なるモダリティの情報を補助情報として用いて予測する。提案手法では、属性の現れやすさの度合いを考慮した画像や属性、ラベルの関係を記述した生成モデルを用いる。複数の評価実験により、提案手法の有効性を示した。

8章では、深層生成モデルを実装するライブラリについて記述している。深層学習のさまざまなライブラリはこれまでも開発されているが、本研究で行うようなマルチモーダル情報に対する深層生成モデルを実装するには適していない。そこで、本研究では、深層生成モデルに特化したライブラリ **Tars** を新規に開発し、提案した。このライブラリでは、確率分布がニューラルネットワークを隠蔽しており、サンプリングや尤度計算が可能である。他にもいくつか深層生成モデルを扱うライブラリがあるが、**Tars** はそのなかでも最も初期のもののひとつである。

本論文は、さまざまな提案手法と評価実験を通じ、近年注目を集める深層学習とマルチモーダル学習の融合を総合的に行っている。それぞれの研究で新規性のある手法を提案しており、その有効性を示すとともに、マルチモーダル学習に対する深層生成モデルのアプローチとして俯瞰的・網羅的なものとなっており、世界的に見てもこういった試みはまだほとんど行われていない。深層学習は産業界でも大きな注目を集めており、本研究のようなマルチモーダルな情報に対する基礎的な研究は、今後の技術の進展に重要である。以上、複数のモダリティをもつ情報に対する深層生成モデルに関して、新規性が高く、また有用な知見を提供しており、その意義は高く評価できる。

よって本論文は博士（工学）の学位請求論文として合格と認められる。