

論文の内容の要旨

論文題目 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

氏名 鈴木 雅大

我々は日々、多くの情報に接している。こうした情報は様々な種類（モダリティ）をとり、我々は複数のモダリティ、すなわち、マルチモーダルな情報を常に取り入れることで、確実な情報処理を行っている。近年は、コンピュータが人間のように複数のモダリティ情報を活用することで、より正確に予測や判断、推論等を行う試みが重要視されている。このような試みは、機械学習の問題と考えることができ、特にマルチモーダル学習と呼ばれる。

マルチモーダル学習の様々な問題設定に共通する困難な点として、モダリティ間の異種性が挙げられる。異種性とは、画像と音声のように互いに形式が大きく異なる性質のことである。さらに、異種性を持つ2つのデータ集合には、特徴空間の違い（各モダリティで表されるデータの形は、次元や構造が大きく異なるということ）と分布の違い（異なるモダリティは、同じ対象を表しているとしても、各モーダル集合の分布には大きな差があるということ）がある。このうち、1つ目の特徴空間の違いへの対処としては、近年、深層ニューラルネットワークを用いた機械学習技術である深層学習によるアプローチが主流となっている。一方で、2つ目の分布の違いについては、深層ニューラルネットワークで直接対処することはできないが、確率的生成モデルによってモダリティ間の分布の違いを明示的に考慮することができる。したがって本論文では、深層学習と生成モデルを組み合わせることで、異なるモダリティ間の異種性の問題に対処できることに着目する。

深層学習と生成モデルを組み合わせるアプローチは大きく分けて2つある。そのうち

の1つは生成モデルの確率分布をニューラルネットワークで表現するアプローチであり、深層生成モデルと呼ばれる。しかし、マルチモーダル学習の問題設定のそれぞれについて、深層学習と生成モデルのアプローチが有効であることを検証した研究はほとんどない。その理由として、深層学習と生成モデルを組み合わせた研究が本格的に行われるようになってから未だ年月が経っていないこと、そして深層生成モデルを実装するための枠組みが十分に整備されていないことが挙げられる。

そこで本論文では、マルチモーダル学習のそれぞれの問題設定（表現・変換・融合・共学習）において、深層学習と生成モデルを利用したモデルを提案して、有効性を検証することを目標とする。さらに、マルチモーダル情報を含んだ深層生成モデルを、実装・利用するためのライブラリを開発することを目指す。

1つ目の研究では、マルチモーダル学習における表現と変換の2つの問題設定に取り組む深層生成モデルを提案する。この研究では、まず従来の表現問題に対するアプローチを単純に深層生成モデルに拡張した joint multimodal variational autoencoder (JMVAE) を検証し、双方向に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や変換したモダリティが崩れてしまい、従来知られていた欠損値補完の手法では解決できないことを示す。この問題を解決するために、本研究では JMVAE-k1 と階層的 JMVAE という手法を提案する。実験から、これらの手法によって単一モダリティを入力とした場合でも適切な共有表現が推論できること、従来の1方向のモデルと比較して同等以上の精度で双方向の変換ができることを示す。

2つ目の研究では、融合、すなわち、複数のモダリティから目標ラベルを予測する問題に取り組む。本研究では、異なるモーダルデータがセットで手に入りやすいのに対して、ラベル情報の獲得は人的コストがかかるという背景から、半教師あり学習に取り組む。半教師あり学習とは、少量のラベルあり集合の他に大量のラベルなし集合がある状況で、汎化性能の高い識別モデルを学習する枠組みである。この章では深層生成モデルを用いた半教師ありマルチモーダル学習の手法として、新たに semi-supervised MVAE (SS-MVAE) と semi-supervised HMVAE (SS-HMVAE) を提案する。また半教師ありマルチモーダル学習では、テスト集合に単一のモダリティしか与えられない設定があることから、モダリティが欠損しても精度を落とさずに目標ラベルを予測する手法として、SS-HMVAE を拡張した SS-HMVAE-k1 というモデルを提案する。これらの手法でモダリティの欠損が補完できること、そして単一モダリティ及びマルチモーダルにおける既存の半教師あり学習と比較して、提案手法の精度が高いことを実験で確認する。

3つ目の研究では、共学習の問題設定の一つとして知られるゼロショット学習について取り組む。ゼロショット学習とは、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枠組みである。本研究では、分類するモダリティとして画像、補助情報として属性を用いる属性ベースゼロショット学習に着目する。属性ベースゼロショット

学習の既存研究では、属性の画像に対する分布の違い（現れやすさ）を明示的に考慮していなかった。本研究では、このような現れやすさの度合いを属性ごとの観測確率と呼び、観測確率を含めて画像や属性、ラベルの関係を記述した生成モデルを提案する。実験では、深層ニューラルネットワークから得た特徴量を利用し、モデルの妥当性の検証及び既存研究との比較実験によって、提案手法が既存手法と比較して有効性の高いモデルであることを示す。

4つ目の研究では、深層生成モデルを実装するライブラリを提案する。深層ニューラルネットワークを用いた生成モデルを記述できるライブラリは既に存在しているが、それらは深層ニューラルネットワークと確率変数を同じレイヤーとして考えており、本研究で提案するようなマルチモーダル情報を持った複雑な深層生成モデルを実装するためには、モデルやネットワークが変更される度に新たに1から実装を行う必要があった。また、深層ニューラルネットワークによって定義された確率分布からサンプリングしたり尤度を計算したりすることは、従来の言語では困難であった。今日の深層生成モデルでは、目的関数として尤度を計算し、潜在変数や各モダリティのデータをサンプリング（推論や生成）することが重要になっている。こうした背景から、深層生成モデルに特化したライブラリ「Tars」を新たに提案する。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、確率分布のそれぞれでサンプリングや尤度計算できる。また、Tarsを用いて様々な分布の深層生成モデルが学習できることを確認する。さらに本ライブラリを使用したアプリケーションを紹介し、本ライブラリで学習した確率分布をそのまま読み込んで画像を生成したりできることを説明する。

以上の研究を通じ、マルチモーダル学習の各問題設定において、深層学習と生成モデルを利用したモデルが有効であることを確認した。また、本論文で提案した深層生成モデルライブラリTarsが、マルチモーダル情報を含むような複雑な深層生成モデルの実装・利用に適していることを確認した。

最後に、本論文の貢献や限界を確認し、今後への課題についてまとめる。そして本研究の産業応用の可能性について触れ、本研究の統一構想や汎用人工知能に向けた考察について述べる。