

令和 2 年 6 月 15 日現在

機関番号：12601

研究種目：研究活動スタート支援

研究期間：2018～2019

課題番号：18H06458・19K21527

研究課題名（和文）モダリティごとの不確実性を考慮した共有表現学習の研究

研究課題名（英文）A study on shared representation learning considering the uncertainty of each modality

研究代表者

鈴木 雅大（Suzuki, Masahiro）

東京大学・大学院工学系研究科（工学部）・特任研究員

研究者番号：30823885

交付決定額（研究期間全体）：（直接経費） 2,300,000円

研究成果の概要（和文）：本研究課題では、画像や文書、音声などの複数の異なる種類の情報（マルチモーダル情報）を統合する方法について取り組む。従来の研究では、モダリティごとの不確実性の違いを考慮しておらず、決定論的に統合していた。本研究では、深層生成モデルと呼ばれる枠組みに基づき、異なるモダリティを確率的に統合することを提案し、複数のマルチモーダル学習の問題設定において、このアプローチが有効であることを示した。また、マルチモーダル情報の関係を含んだ複雑な深層生成モデルを簡潔に実装するため、新たにライブラリを開発した。これらの成果は国内論文誌や国際会議などで発表した。

研究成果の学術的意義や社会的意義

本研究で提案する異なるモダリティの統合の枠組みは、今回扱ったデータや問題設定によらず、様々な領域に応用できると考えている。それは、この統合方法では深層生成モデルを用いているため、モダリティの不確実性の違いのみに着目しており、モダリティの入力空間の次元数には依存しないからである。また、今回開発した深層生成モデルライブラリは、マルチモーダル学習のモデルに限らず、様々な深層生成モデルの実装に利用することができる。

研究成果の概要（英文）：In this research, we addressed how to integrate several different types of information (i.e., different modalities), such as images, documents, and sounds. Previous studies did not take into account the differences in uncertainty across modalities and therefore integrated them deterministically. In this study, we proposed the probabilistic integration of different modalities based on a framework called deep generative models. We then showed that this approach is effective in multiple multimodal learning problem settings. In addition, we developed a new library to simplify the implementation of complex deep generative models containing multimodal information relationships.

研究分野：人工知能

キーワード：深層学習 共有表現学習 マルチモーダル学習 深層生成モデル

1. 研究開始当初の背景

コンピュータが扱う情報の種類は、初期の数値から始まり、近年では画像、音声、文書、動画と多様化してきている。こうした異なる種類の情報はモダリティと呼ばれ、人工知能分野では、画像認識、音声認識、自然言語処理といったように、モダリティごとに異なるタスクとして独立に研究が進められてきた。

近年、深層学習の技術的發展により、いずれのタスクについても飛躍的に認識精度が向上している。例えば、画像認識については2015年の時点で人間の認識精度を上回っており、音声認識についても2017年時点で人間とほぼ同様の認識精度を達成したと報告されている。

深層学習の發展がもたらした恩恵は、個々のタスクの精度向上だけではない。深層学習による手法では、モダリティに応じて異なる種類のモデル(畳み込みネットワークや再帰結合型ネットワークなど)が使われるものの、いずれも階層構造を持った深層ニューラルネットワーク(以下DNN)で構成される。そのため、同一の最適化手法を用いてend-to-endに学習することができる。このことから、異なるモダリティを統一的に扱うマルチモーダル学習の研究が数多く行われるようになった。

マルチモーダル学習の問題設定は様々考えられるが、いずれの問題についても、どのように異なるモダリティを統合するかという問題に帰着する。従来は、各モダリティ同士が一对一対応するように(決定論的に)学習することが主流であった。しかし、通常各モダリティの関係性は決定論的に対応させることができない。例として、画像とタグ情報という2つのモダリティを考えよう。海が写っている画像があった場合、「海」というタグをほぼ一意に割り当てることができるだろう(図1(a))。しかし、逆に「海」というタグ情報から想定される画像は無数に存在する(図1(b))。このようにモダリティの種類によって情報量が異なるため、その対応関係は非対称になる。したがって、従来の決定論的な学習方法では適切にモダリティの統合ができない可能性が考えられる。

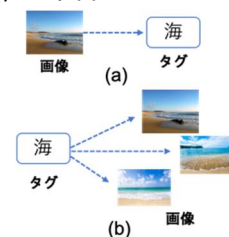


図1: 異なるモダリティの関係の非対称性

2. 研究の目的

そこで本研究では、異なるモダリティを適切に統合するために、深層生成モデルを用いることを提案する。深層生成モデルとは、データの関係性に基づく確率分布によって明示的にモデル化するアプローチである。これに基づき、異なるモダリティの「不確実性の違い(情報量の違い)」を考慮した、確率的な統合を実現することを目指す。

本研究の目的は、この深層生成モデルによるアプローチが、マルチモーダル学習において有効であることを示すことである。

3. 研究の方法

本研究では、目的を達成するために、以下の2つのサブ研究を進めた。

1つ目は、提案するアプローチを、様々な問題設定で検証することである。それぞれの問題設定に応じて、深層生成モデルに基づくマルチモーダル学習の手法を提案し、実験によってその有効性を確認した。

2つ目は、それらを実装するためのライブラリの開発である。今回提案するアプローチでは、DNNに基づく確率分布の数が、モダリティの数やモデルの複雑さに応じて増大する。したがって、こうした複雑な深層生成モデルを簡潔に実装できるライブラリが重要となる。本研究では、提案アプローチを実現するための独自のライブラリを開発し、その有効性を検証した。

4. 研究成果

(1) 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

サブ研究1の成果の1つとして、深層生成モデルを用いた半教師あり学習について説明する。

マルチモーダル学習の問題設定の一つとして、複数のモダリティを入力とし、対応するラベルを予測するという「融合(fusion)」がある。通常、これを実現するためには大量のラベルありデータが必要であるが、ラベル付けのための人的コストがかかるという課題がある。これを解決するために、ラベルがないマルチモーダルデータを用いることで汎化性能を向上させる「半教師ありマルチモーダル学習」を考える。本研究では、深層生成モデルを用いた半教師ありマルチモ

ーダル学習モデルを提案した。

表1に、手書き数字画像データセット MNIST を加工しマルチモーダルデータセットとして(2通りの異なるノイズを付与して、それぞれを異なるモダリティとみなす)学習した場合の結果を示す。

提案手法は、異なるモダリティの不確実性を考慮した(確率的に統合した)ものである。比較手法として、モダリティを単純に統合して既存の半教師あり学習[1]を適用した場合の結果を載せている。既存手法がラベルありデータ数が少なくなると正解率が大幅に下がるのに対して、提案手法は高い正解率を維持していることがわかる。この結果から、半教師ありマルチモーダル学習の問題設定において、提案アプローチの有効性を確認することができた。

モデル	ラベルありデータ数		
	500	300	100
既存手法[1]	89.03	88.85	70.68
提案手法	93.86	93.71	90.80

表1: 既存手法(マルチモーダルデータを単純に結合して半教師あり学習手法[1]を適用)と提案手法の、ラベルありデータ数を変えたときの正解率(%) (元論文の表から再構成。パラメータ等は元論文参照)

この研究以外にも、画像の表現をラベル情報という別のモダリティを利用してうまく分離するように学習する研究や、与えられた服画像を利用して人物画像の服を着せ替える研究などを行った。また、深層生成モデル(およびDNNで設計されていない通常の確率的生成モデル)に基づき、人間のように複数のモダリティを統合可能な大規模な認知アーキテクチャの枠組みを提案した。

これらの結果から、提案アプローチの有効性を検証できたと考えている。

(2) Pixyz: 複雑な深層生成モデル開発のためのフレームワーク

(1)のようなモデルを実装するために、サブ研究2の成果として、Pixyzという深層生成モデルライブラリを開発した。

Pixyzは、深層生成モデルの特徴である、1. 確率分布によるネットワークの隠蔽、および2. 複数の項で構成される目的関数、という2つの特徴を念頭に、独自のAPIによって構成されている。これによって、従来のライブラリでは実装困難だった、複数のモダリティを含んだ複雑な深層生成モデルでも容易に実装することができる。

また、シンプルな構成となっているため、特に深層生成モデルを実装する場合において、実行速度が従来の確率モデリングライブラリよりも高速となっている。表2は、Pixyzと、既存の確率モデリングライブラリであるPyro[2]との実行速度の比較結果である。これは、深層生成モデルの一つであるvariational autoencoder(VAE)をMNISTで学習したものである。比較結果より、Pixyzの方がより高速で学習できることがわかる。

潜在変数のサイズ	隠れ層のサイズ	Pyro[2] (ms)	Pixyz (提案ライブラリ) (ms)
10	400	4.91±0.12	3.61±0.11
30	400	4.94±0.13	3.58±0.10
10	2000	4.93±0.12	3.62±0.09
30	2000	4.95±0.12	3.65±0.09

表2: 異なるサイズのVAEにおけるPyroとPixyzのステップあたりの実行時間の比較(元論文の表から再構成。パラメータ等は元論文参照)

したがって、速度面において、提案ライブラリの有効性を示すことができた。また、(1)において実際に本サブ研究で開発したライブラリを利用していることから、本研究テーマにおいて、このライブラリが有効であることが確認できた。

以上より、本研究の目的を予定通り達成することができた。これらの結果は、国内論文誌や国際会議、国内学会にて発表した。

< 引用文献 >

[1] Kingma, D. P., Mohamed, S., Rezende, D. J. and Welling, M.: Semi-supervised learning with deep generative models, Advances in Neural Information Processing Systems, pp. 3581-3589 (2014).

[2] Bingham, E., Chen, J. P., Jankowiak, M., Obermeyer, F., Pradhan, N., Karaletsos, T., Singh, R., Szerlip, P., Horsfall, P., and Goodman, N. D.: Pyro: Deep universal probabilistic programming, arXiv preprint arXiv:1810.09538 (2018)

5. 主な発表論文等

〔雑誌論文〕 計3件（うち査読付論文 3件/うち国際共著 0件/うちオープンアクセス 0件）

1. 著者名 鈴木雅大, 松尾豊	4. 巻 59
2. 論文標題 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	5. 発行年 2018年
3. 雑誌名 情報処理学会論文誌	6. 最初と最後の頁 2261-2278
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 久保静真, 岩澤有祐, 鈴木雅大, 松尾豊	4. 巻 60
2. 論文標題 服の領域を考慮した写真上の人物の自動着せ替えに関する研究	5. 発行年 2019年
3. 雑誌名 情報処理学会論文誌	6. 最初と最後の頁 870-879
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Taniguchi Tadahiro, Nakamura Tomoaki, Suzuki Masahiro, Kuniyasu Ryo, Hayashi Kaede, Taniguchi Akira, Horii Takato, Nagai Takayuki	4. 巻 38
2. 論文標題 Neuro-SERKET: Development of Integrative Cognitive System Through the Composition of Deep Probabilistic Generative Models	5. 発行年 2020年
3. 雑誌名 New Generation Computing	6. 最初と最後の頁 23 ~ 48
掲載論文のDOI (デジタルオブジェクト識別子) https://doi.org/10.1007/s00354-019-00084-w	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計7件（うち招待講演 1件/うち国際学会 3件）

1. 発表者名 鈴木雅大, 松尾豊
2. 発表標題 半教師ありマルチモーダル深層生成モデルにおける共有表現の有効性と単一モダリティ入力への拡張
3. 学会等名 2018年度人工知能学会全国大会
4. 発表年 2018年

1. 発表者名 Hirono Okamoto, Masahiro Suzuki, Itto Higuchi, Shohei Ohsawa, Yutaka Matsuo
2. 発表標題 Dual Space Learning with Variational Autoencoders
3. 学会等名 Workshop on Deep Generative Models for Highly Structured Data, International Conference on Learning Representation (国際学会)
4. 発表年 2019年

1. 発表者名 鈴木 雅大, 金子 貴輝, 谷口 尚平, 松嶋 達也, 松尾 豊
2. 発表標題 Pixyz: 複雑な深層生成モデル開発のためのフレームワーク
3. 学会等名 2019年度人工知能学会全国大会
4. 発表年 2019年

1. 発表者名 久保 静真, 岩澤 有祐, 鈴木 雅大, 松尾 豊
2. 発表標題 身体の3次元構造を考慮したニューラル仮想試着
3. 学会等名 2019年度人工知能学会全国大会
4. 発表年 2019年

1. 発表者名 Masahiro Suzuki
2. 発表標題 Pixyz: a framework for developing complex deep generative models
3. 学会等名 Workshop on Deep Probabilistic Generative Models for Cognitive Architecture in Robotics (IROS2019) (招待講演) (国際学会)
4. 発表年 2019年

1. 発表者名 鈴木雅大
2. 発表標題 深層生成モデルと世界モデル
3. 学会等名 第4回統計・機械学習若手シンポジウム
4. 発表年 2019年

1. 発表者名 Shizuma Kubo, Yusuke Iwasawa, Masahiro Suzuki, Yutaka Matsuo
2. 発表標題 UVTON: UV Mapping to Consider the 3D Structure of a Human in Image-Based Virtual Try-On Network
3. 学会等名 Workshop on Computer Vision for Fashion, Art and Design, The IEEE International Conference on Computer Vision (ICCV 2019) (国際学会)
4. 発表年 2019年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考