

研究終了報告書

「異質データ間での深層転移学習の探求」

研究期間： 2021 年 10 月～ 2024 年 3 月

研究者： 森田 堯

1. 研究のねらい

動物音声とヒトの言語との比較研究を行う場合、人為的に抽出した言語の特徴的要素が動物音声においても見られるか否かを調査する形式をとる(例:単語のような構成単位の出現頻度が Zipf 分布に従うか否か)。しかし、人為的な特徴要素の選出は現行の言語理論や生物学的知見に依存するため、言語・動物音声間に未知の共通要素が存在した場合には類似性を見落とすことになる。このような「研究者に取りこぼされた知見」を補完するためには、AI 技術を活用しデータ駆動的に言語と動物音声の類似度を評価する方法が有効である。

データ駆動的な類似度評価の方法として、研究代表者は人工ニューラルネットワークの転移学習を用いる発想に至った。転移学習は大量入手が可能なデータを用いて予備的にニューラルネットワークを訓練し、得られたネットワークの一部分のみを主目的のデータで再訓練するという技術であり、特に大規模なデータ収集が困難なデータに深層学習を応用する際に用いられる。転移学習が効果的に動作するためには事前学習で用いられる大規模データと転移先のデータとの間に共通の性質が存在している必要があるため、典型的な転移学習では転移先のデータとの類似性が明らかなデータが事前学習に用いられる。しかし、転移学習に有効なデータ間の類似度は人間の目で見ても明らかなものだけであるとは限らず、最近の研究ではピアノ演奏やコンピュータプログラムのコードが自然言語処理の事前学習に一定の効果を発揮するという結果が報告されている(Papadimitriou & Jurafsky, 2020, *EMNLP*)。異質データ間で転移学習が成り立つことは、事前学習用データと転移先データの間に潜在的な類似性が存在することを示唆するため、転移学習を用いた言語と動物音声の比較を始めとする、異質データ間での転移学習を探求し、転移可能性を尺度とする未知の類似性を発見することが本研究のねらいであった。

2. 研究成果

(1) 概要

当初の研究計画に従い、動物音声からヒトの言語への転移学習を実施した。動物音声の予測学習を行った回帰型ニューラルネットワークを、言語モデリング課題(逐次的な単語予測課題)に転移し、転移先課題の精度によって、両データに潜在する時系列パターンの類似度を評価した。ジュウシマツ及びテナガザルの音声から、Wiki-40B データセットで公開されている 41 種類のテキスト言語データへの転移学習を網羅的に検証した。

また、動物音声からヒトの言語への転移学習の前処理技術として、テナガザル音声の音源定位技術の開発に取り組んだ。今回使用したテナガザル音声は、3 頭が同居する飼育室内で録音されたため、個体識別ができていなかった。飼育室内において、テナガザルは個別のケージ内で飼育されており、録音された音声から「どこで発せられたのか」を推定することで、「誰が発声したのか」を推定することが可能となる。そのため、多チャンネルマイクロフォンアレイによる録音データに基づく音源定位を探求した。

加えて、Transformer が時間表現として採用している位置エンコーディングが、通常内部演

算によって時間表現を行う回帰型ニューラルネットワークの性能を向上させるか否かの検証を行った。ランダム生成した記号列を逆順で再生成する回文生成課題に基づく検証の結果、特に記号数が大きい(語彙が大きい)場合に位置エンコーディングによって回帰型ニューラルネットワークの性能が向上することを発見した。回帰型ニューラルネットワークにとって理論上「余分な」データが記憶領域を圧迫しているに関わらず、記憶課題の性能が向上するという本研究の知見は、回帰型ニューラルネットワークに関する今後の研究開発に対する重要な示唆となった。

(2) 詳細

研究テーマ A「深層転移学習による動物音声とヒトの言語の類似度計測」

本研究の主眼である、動物音声とヒトの言語との間の潜在的類似度を、転移学習によって計測した。研究手法としては、持続的に発せられた動物音声を事前学習、ヒトの言語(単語列)を転移先に用いる転移学習を研究した。入力埋め込み層、時系列処理層、出力層の 3 つのモジュールからなる人工ニューラルネットワークを使用し、動物音声を用いて全体を事前学習した後、言語データを用いて時系列処理層以外を再学習させた。すなわち、表面的な記号・値の差異は無視した上で、潜在的な時系列パターンがどの程度転移可能かを調査した。

事前学習用データには、動物 2 種(ジュウシマツ、アジルテナガザル)の音声に加え、比較対象としてピアノ演奏(MAESTRO データセット; Hawthorne et al., 2019, *ICLR*)、及び階層的言語モデル(確率的文脈自由文法; PCFG)から生成した人工データを検証した。転移先の言語データには、Wiki-40B データセット(Guo et al., 2020, *LREC*)で公開されている 41 種類のテキスト言語データを採用し、各事前学習データから各言語への転移学習成績を網羅的に検証した。

研究テーマ B「動物音声用音源定位技術の開発」

研究テーマ A で用いたジュウシマツの音声は、各個体を防音室に隔離した上で録音されており、個体識別が完了している。一方、テナガザルの音声は、3 頭が同居する飼育室内で録音されたため、発声個体の同定ができていなかった。個体識別を行うことで、音声分類時における個体差を除去することができ(Morita et al., 2021, *PLOS Comp. Biol.*)、時系列分析においても影響が生じる。飼育室内において、テナガザルは個別のケージ内で飼育されており、録音された音声は「どこで発せられたのか」を推定することで、「誰が発声したのか」を推定することが可能となる。そのため、多チャンネルマイクロフォンアレイによる録音データに基づく音源定位を探索した。

研究テーマ C「回帰型ニューラルネットワークの時間表現に関する研究」

研究テーマ A の実施過程において、近年の言語モデルで用いられる Transformer (Vaswani et al., 2017, *NeurIPS*)が用いる時間表現(位置エンコーディング)が単体で極めて強力であるという示唆を得た。この結果を基に、回帰型ニューラルネットワークにおける位置エンコーディングの有効性について検証した。

Transformer は構造上フィードフォワード型ニューラルネットワークの 1 種であり、入力データ(トークン)間の時系列関係を内部演算によって表現することができない。そのため、各トークンの出現時刻を表す「タイムスタンプ」を付与する位置エンコーディングによって、言語を始めとする時系列データを処理する。

これに対し、回帰型ニューラルネットワークはトークンを逐次的に読み込むため、トークン間の時系列関係を内部演算によって表現することが、原理上可能である。そのため、位置エンコーディングを、回帰型ニューラルネットワークの入力に対して「冗長的に」付与する検証は、本研究代表者の認知する範囲では行われていなかった。

本研究では、位置エンコーディングが回帰型ニューラルネットワークの能力拡張にも寄与するのではないかという仮説の下、特に大規模な語彙と記憶力に関して実験的検証を行った。ランダム記号列を記憶し、その出現順序を反転させた記号列を出力する回文生成課題を回帰型ニューラルネットワークに学習させ、回文の正答率によって位置エンコーディングの有効性を検証した(図 1)。この回文課題は、「観測データを順に記憶領域に追加

し、最新のデータから順に記憶領域から取り出す」スタック処理によって実行可能であり、回帰型ニューラルネットワークへの入力に対する位置エンコーディングは理論上不要である。位置エンコーディングの実装は、多様な周波数の正弦・余弦波を組み合わせる標準的なものを採用した(Vaswani et al., 2017, *NeurIPS*)。

3 種類の代表的な実装(Elman・GRU・LSTM)を用いた検証の結果、位置エンコーディングは回帰型ニューラルネットワークの回文出力性能を向上させることが判明した(図 2)。特に語彙が大規模化するにつれ、位置エンコーディングなしのネットワーク(薄色)の正答率が低迷したのに対し、位置エンコーディングの付与をした場合(濃色)では順調に学習が進んだ。

回帰型ニューラルネットワークにとって理論上「余分な」データが記憶領域を圧迫しているにも関わらず、記憶課題の性能が向上するという本研究の知見は、回帰型ニューラルネットワークに関する今後の研究開発に対する重要な示唆となった。

領域内研究者間連携

本領域での研究者間連携の成果として、網野海氏(東京大学)と実施予定(2024 年 2 月-3 月開催で検討中)のシンポジウム「考える AI と共存する生物研究者(仮称)」(主催:進化学若手の会)が挙げられる。

動物研究に対する機械学習の応用研究は近年数多く報告されているが、研究対象の生物種に応じた個々の研究が点在している状況であり、汎く動物研究における機械学習技術に関する情報交換・議論の場は限られている。実際、蝶を中心とした昆虫の研究を行っている網野

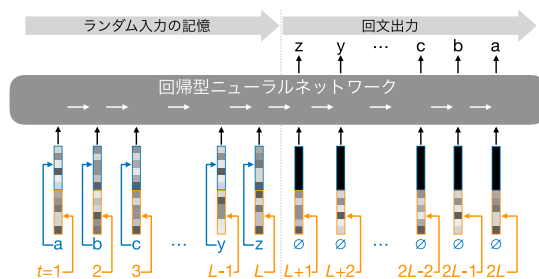


図 1: 回文生成課題による回帰型ニューラルネットワークに対する位置エンコーディング(橙)の有効性検証。位置エンコーディングは単語埋め込み(青)に結合して入力。

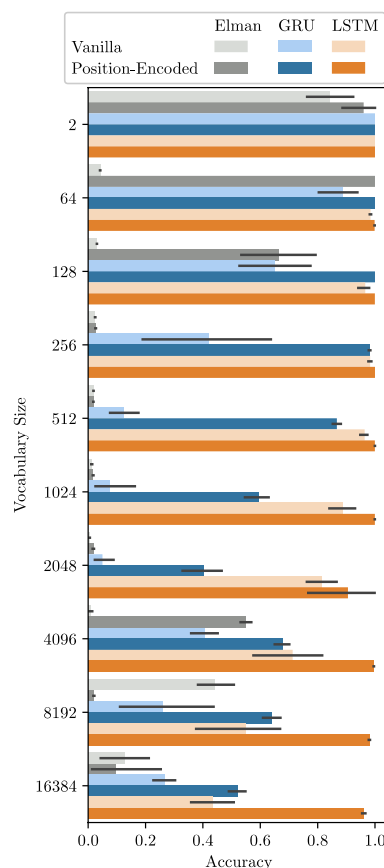


図 2: 位置エンコーディングの有無による回文生成課題の正答率。

氏と、霊長類を中心に研究している本研究代表者との交流は、本領域以前には皆無であった。そのため、上記シンポジウムは多様な研究対象を持つ動物研究者が、種間横断的に情報交換・議論を行うための貴重な機会であり、その開催の起点となった本領域の存在意義は大きい。

3. 今後の展開

本研究を含め、研究提案者は、これまで、動物が自発的に発した音声や動物の自発的移動の軌跡データに対し、機械学習技術を活用した分析を行ってきた。動物の自発的行動は、自動観測による大規模データ収集が可能であるため、機械学習との親和性が高いが、自発的行動の分析から得られる知見は、あくまで「動物が通常何を行っているか?」という自然状態に関することに限定され、「動物にどのようなことができるか?」という潜在能力に関する問いには答えることが出来ない。

今後は、自発行動データの分析に留まらず、実験研究と機械学習との融合により、これまで得られなかった新たなデータを引き出す研究に取り組む計画である。

4. 自己評価

研究目的の達成状況

「深層転移学習によって、動物の音声とヒトの言語の類似度を測定する」という目的自体は、地道な計算によって概ね達成することができた。しかし、追加検証が難航し、論文として発表可能な成果としてはまとまっていない。したがって、本研究の目的に対する満足な達成状況が得られたとは言い難い。

研究の進め方(研究実施体制及び研究費執行状況)

研究実施体制については、既存のデータの活用したため、新規データ収集に必要な研究者間連携は大きくなく、研究代表者の個人研究課題として実施する上での問題はなかった。

研究費の執行状況についても、研究課題に対して過不足なく執行できており、問題はなかった。

研究成果の科学技術及び社会・経済への波及効果

本研究期間中においては、1 件の国際会議発表(査読付き論文あり)と 1 件の招待講演を通じて、本研究を含む動物研究への機械学習活用の取り組みを周知することができた。特に、「画像処理技術を応用して動物を動画から抜き取る」といった自明な応用例以外の機械学習活用可能性を周知できたことは、自身としても評価できる。

その一方、本研究の主たる成果を期間中に論文や国際会議で発表することができなかった点は問題として認識している。しかし、追加解析を今後進めた上で、研究成果を論文にすることができれば、データ駆動型の比較動物研究の先駆けとして学術的インパクトを残せると見込んでいる。

5. 主な研究成果リスト

(1) 代表的な論文(原著論文)発表

研究期間累積件数: 1件

1. “On the relation between context dependency and vocabulary in human language and birdsong”, *Proceedings of the Joint Conference on Language Evolution 2022*, pp. 517–524. 2022. T. Morita* (1/4), H. Koda. 査読あり.

(2) 特許出願

研究期間全出願件数: 0 件 (特許公開前のものは件数にのみ含む)

(3) その他の成果 (主要な学会発表、受賞、著作物、プレスリリース等)

1. “動物音声研究のための機械学習”, 第 57 回関西合同音声ゼミ, 2023 年 7 月 8 日, 大阪府吹田市・茨木市. 招待講演.
2. “Towards Biological Discoveries through the Eyes of AI”, 室蘭工業大学異分野研究セミナー, 2022 年 5 月 25 日, 北海道室蘭市. 招待講演.
3. “認知・行動科学のための機械学習”, 日本音響学会 2022 年春季研究発表会, 2022 年 3 月 9 日, オンライン開催. 招待講演.