探査子法を用いた音楽から学習可能な言語モデルの構文的性質の解析

加藤 万理子 [1]　高橋 信行 [1]
[1] 公立はこだて未来大学
{b1019090,ntakahas}@fun.ac.jp

概要

自然言語が持つ構造と似た構造を持つ非言語データとして音楽やプログラムコードのデータを事前学習に用いる研究から, 非言語データから自然言語習得に寄与する構造的知識が得られていることが先行研究によりわかっている. 一方で, 音楽データについては構造解釈が複雑なことから具体的にどのような構造が獲得されているか明らかになっていない. 本研究では音楽データに着目し, 音楽データが自然言語習得に寄与することを確認したのち, 探査子を用いて言語モデルが音楽データから学習している構文レベルの言語学的性質を明らかにする.

1　はじめに

事前学習済み言語モデルはさまざまな自然言語タスクで高い実験的パフォーマンスを示している [1]. こうした進歩により, 言語モデルが学習している言語情報を理解することへの関心が高まっている. また, ある言語で事前学習された言語モデルはその言語内だけでなく, 学習した言語とは異なる言語の推論精度も向上することが先行研究にて示されている [2]. これは, 事前学習において単一の言語に限らない知識が学習されていることを示唆しているが, どのような知識が学習されているのかはまだ十分に解明されていない.

更に, この事前学習で得られている自然言語に関する知識は, 非言語データからも得られることがわかっている. 音楽やプログラムコードのデータを事前学習に用いた研究 [3] では, 事前学習により自然言語に関する構造的な知識が得られていることを明らかにしており, 比較対象のランダムデータと比較して, 音楽やプログラムコードなどの非言語データは自然言語への転移を行った際の Perplexity スコアが向上することが示されている. しかし音楽データの構造解釈は数多くあり, 具体的にどのような構造が学習されているのかはまだ明らかになっていない.

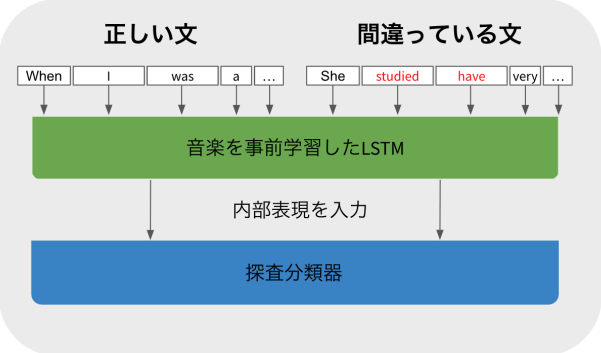

図 1　探査子法の実験プロセス

本研究では言語モデルが音楽データから獲得している知識のうち, 特に自然言語に関する構造的知識が獲得されているのかどうかについて着目した.

まず始めに, Papadimitriou らが提案した Test for Inductive Bias via Language Model Transfer(TILT) [3] を用いて音楽データを用いて事前学習を行い, ファインチューニングで自然言語への転移学習を行う. このときの自然言語への転移性能を調べることで, 自然言語習得に寄与する構造を持っていることを確認する. その後, 探査子という手法を用いて事前学習時の内部表現を調べることで, 言語モデルが音楽データからどのような言語に関する構造的知識を獲得しているかどうかを調べた.

2　アプローチ

本研究では図 1 に示した探査子という手法を用いて, 音楽データを事前学習したモデルがどのような言語に関する構造的知識を獲得しているのかを調べた. また, Papadimitriou らが提案した Test for Inductive Bias via Language Model Transfer(TILT) [3] フレームワークを用いて音楽データの事前学習を行なった. 具体的には以下のようなステップで実験を

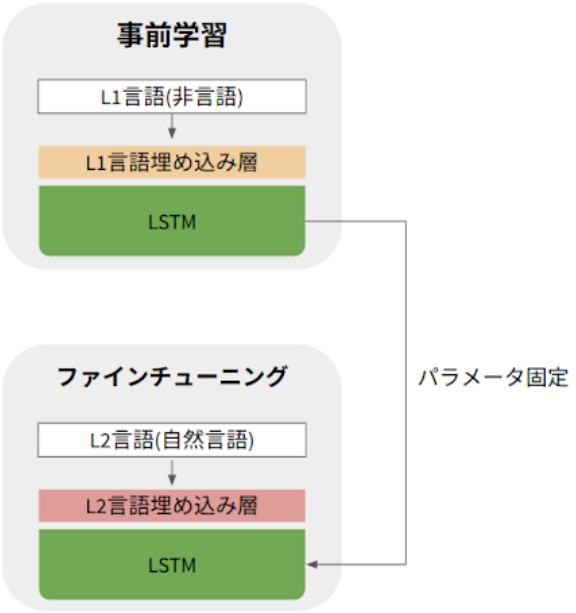

図 2　TILT フレームワークでのプロセス

行なった.

1. TILT フレームワークを用いて, 非言語データで言語モデルを事前学習する
2. 同フレームワークにて更に英語でファインチューニングを行う
3. ファインチューニングを行なったモデルに対して文章を入力し, 内部表現を取り出す
4. 内部表現を入力とし, 入力した文章のラベルを予測する探査分類器を学習, 評価する

ステップ 1,2 について 2.1 節, ステップ 3, 4 については 2.2 節にて説明を行う. その後は実験設定について 3 節にて説明を行い, 4 節で実験結果を示す.

2.1　TILT フレームワーク

Papadimitriou らが提案した Test for Inductive Bias via Language Model Transfer(TILT) [3] フレームワークでは言語モデルが事前学習時に得ている情報のうち, 特に構造的な知識について調べることが可能である. 具体的には事前学習に用いられる言語 (L1 言語) の構造が, ファインチューニングに用いられる言語 (L2 言語) に対してどの程度転移可能であるかを調べることができる. 図 2 に TILT フレームワークの学習プロセスを示す. ファインチューニング時はパラメータ更新を行わず埋め込み表現のみ学習することで, 事前学習時に用いる L1 言語の構造のみを学習させることが可能である.

表 1　非言語データの例

MAESTRO [6]	4D 54 68 64 00 00 00 06 00 01 ...
Zipf 分布 [3]	en con conocidas y en los victoriano como trabajar hunki monte * en juegos días en el
一様分布 [3]	marroquín jemer pertenecer osasuna formaron citoesqueleto relativismo

2.2　探査子

探査子の目的は, モデルの内部表現が実際にどのような情報を獲得しているかを明らかにすることである. 探査子実験で行う学習プロセスを図 1 に示す. この手法では, 事前学習済み言語モデルの内部表現を分類器の入力として与え, 対象とする言語学的性質を予測する. ここで分類器を使って学習するタスクを探査タスクという. 本研究では SentEval ライブラリ [4] を用いて, バイグラムシフトタスク, 木構造深さタスク, 最上位構成要素タスクの 3 つを [5] 行う. 以下各タスクについて詳説する.

バイグラムシフトタスク　二値分類タスクであり, 正しい語順であるかどうかの判定を行う. 探査分類器モデルは正しい語順である文と, ランダムに隣り合う 2 つの単語を反転させた文を区別するよう学習される.

木構造深さタスク　文の階層構造を分類する多値分類タスクである. 特に, ルートから任意の葉までの最長パスの深さによって文の分類を行う. ここで, 木構造の深さは文の長さに相関していることから, 深さの範囲が 5 から 12 であるようにサンプリングを行うことで調整を行う.

最上位構成要素タスク　20 クラスを分類するタスクであり, 文ノードの直下にある最も上位である構成要素の品詞タグについて分類する.

3　実験設定

3.1　データセット

本研究では, 音楽と言語の他に, ベースラインとしてランダムデータを用いて実験を行なった. 音楽データとしては, Papadimitriou らが用いた [3] クラシックのピアノ演奏が含まれる MASETRO データセット [6] を用いた. 言語データは Wikipedia から英語データを作成した. ベースラインとして一様分布と Zipf 分布を用いて英語データからサンプリングを行い, ランダムデータを生成した. Zipf 分布は自然言

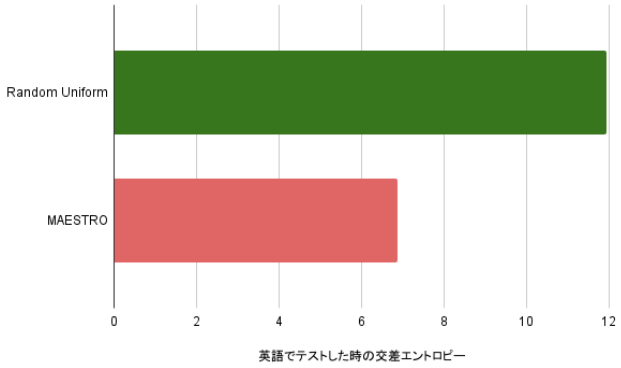

図 3　ファインチューニング後の Perplexity スコア

語における語彙分布をよく表しており, 一様分布からサンプリングしたデータよりも高い性能を示すことが予想される. しかしこれらの分布は自然言語の文法やかかり受けなど構造に起因する要素を反映していない. よって, 非言語データがランダムデータよりも高い転移性能を示した場合, 言語モデルが非言語データから自然言語に関する構造的知識を得ているという解釈が可能である.

3.2　探査分類器

探査実験では, 音楽データを事前学習した LSTM を用いて, 最後の層の内部表現を入力とし, 探査分類器を学習させる. ここで, 探査分類器自体が言語学的性質を学び, 探査タスクでの評価が向上するのを防ぐため, 探査分類器に用いるモデルのアーキテクチャはシンプルなほうが望ましいことから, 探査分類器は線形モデルと MLP モデルを用いた.

4　実験結果

4.1　TILT フレームワークを用いた事前学習

TILT フレームワークを用いて非言語データから英語への転移性能 (Perplexity スコア) を調べ, 現在結果が出ている一様分布のランダムデータと音楽データでの結果を図 3 に示す. MAESTRO データセットについて, ベースラインであるランダムデータよりも自然言語への転移性能が良いことを確認できる. この結果から音楽データから自然言語へ転移可能な構造的知識が学習されていることが確認できた. Zipf 分布に関するシミュレーションの一例では, 音楽データの転移性能が優れていることが示唆されたが, この有意性を確認するための統計処理を行うために, 現在複数のシミュレーションを行っている.

4.2　探査実験

バイグラムシフトタスクにおいて, Zipf 分布と一様分布で作成したランダムデータを事前学習に用いたモデルでは, 精度は 0.7 前後の値に留まることが確認できた. 一方音楽データを事前学習に用いたモデルでも, 精度は 0.7 前後の値を記録しており, ランダムデータの場合と有意な差が確認できなかった.

5　関連研究

マルチモーダルな入力を受け取る深層学習モデルを対象とし, 画像とそれに対するキャプションの対応関係を調べる探査タスクに関する先行研究がある [7]. ここでは, キャプションに対して同義語や多義語を用いることで, 深層学習モデルが画像内のオブジェクトとキャプションの対応関係を獲得しているかどうかを調べている.

一方で, マルチモーダルな入力を受け取る深層学習モデルが獲得している知識について, 画像の分野と異なり音楽分野では調査がされていない. この要因の一つとして, 画像と自然言語の対応関係と比較して, 音楽の特徴を自然言語で表すことが難しいことや, 時系列を含むデータであることなどが指摘されている [8].

本研究では音楽データを対象としているが, 音楽データと自然言語の対応関係は不明瞭であることから, 対応関係を探るような探査タスクの構築は困難であると予想される. 一方で, 検討の出発点として, 自然言語に関する情報が獲得されているかを調べることができる既存の探査タスクを用いることで, 音楽データと自然言語の対応関係は不明ではあるものの, 獲得している情報を調べることは可能であると考えられる.

6　おわりに

本研究では音楽データに着目し, 音楽データが自然言語習得に寄与することを確認したのち, 探査子を用いて言語モデルが音楽データから学習している構文レベルの言語学的性質ついて解析を行なった. その結果, 先行研究にて示されていた音楽データから学習された自然言語へ転移可能な知識の存在が確認できた. 一方で, 探査手法を用いてランダムデータと比較実験を行なった結果, 音楽データを事前学習に用いたモデルは入力された文章に対して正しい語順であるか否かの判別は, 今回の実験では有意な差

は見られなかった. 年次大会では未実施の探査タスクを用いた実験を行い, 言語モデルが獲得している言語学的性質について更に分析を実施した結果を発表する予定である.

参考文献

[1] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. Exploring the limits of transfer learning with a unified text-to-text transformer, 2020.

[2] Mikel Artetxe, Sebastian Ruder, and Dani Yogatama. On the cross-lingual transferability of monolingual representations. In **Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics**, July 2020.

[3] Isabel Papadimitriou and Dan Jurafsky. Learning Music Helps You Read: Using transfer to study linguistic structure in language models. In **Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)**, November 2020.

[4] Alexis Conneau and Douwe Kiela. Senteval: An evaluation toolkit for universal sentence representations. **arXiv preprint arXiv:1803.05449**, 2018.

[5] Alexis Conneau, German Kruszewski, Guillaume Lample, Loïc Barrault, and Marco Baroni. What you can cram into a single $&!#* vector: Probing sentence embeddings for linguistic properties. In **Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, July 2018.

[6] Curtis Hawthorne, Andriy Stasyuk, Adam Roberts, Ian Simon, Cheng-Zhi Anna Huang, Sander Dieleman, Erich Elsen, Jesse Engel, and Douglas Eck. Enabling factorized piano music modeling and generation with the maestro dataset. **arXiv preprint arXiv:1810.12247**, 2018.

[7] Adam Dahlgren Lindström, Johanna Björklund, Suna Bensch, and Frank Drewes. Probing multimodal embeddings for linguistic properties: the visual-semantic case. In **Proceedings of the 28th International Conference on Computational Linguistics**, December 2020.

[8] Andrea Agostinelli, Timo I. Denk, Zalán Borsos, Jesse Engel, Mauro Verzetti, Antoine Caillon, Qingqing Huang, Aren Jansen, Adam Roberts, Marco Tagliasacchi, Matt Sharifi, Neil Zeghidour, and Christian Frank. Musiclm: Generating music from text, 2023.