4W-07

「人間らしい」比喩生成における汎用自然言語処理モデルと人間の Attention の比較 ～視線情報をもちいて～

中村 航†　荻原 佑介‡
東京理科大学 経営学部 ビジネスエコノミクス学科

1.本研究の位置付け

機械学習の自然言語処理の領域において, 大規模言語モデルは, テキストデータからのパターン学習により, 言語の構造や意味を高度に把握する能力を備えている. BERT のような先進的なモデルについては, その性能を人間の言語理解の観点から再解釈するといった研究も進んでいる. たとえば, BERT に短歌を入力したときの振る舞いについて, 人間の脳神経活動との対応関係を観測した研究がある[1]. この研究では, BERT の深い層は, 抽象的な情報を捉える脳の部位と強い相関をもち, 文の意味的な性質を捉えている点などを明らかにしている. 脳科学や認知科学の観点からの機械学習モデルの検討は, ブラックボックスの性質を持つ大規模言語モデルの内部構造を明らかにする可能性を秘めている.

しかしながら, これまでの研究では, BERT の性能を飛躍的に向上させた Attention 機構に対して, 人間の生体情報との対応による吟味が行われていない. そこで本研究では,

1)Attention 機構に焦点を当て, 人間の意思決定過程から BERT のタスク処理を評価する. 具体的には, 人間の感情や直感に基づいて生成される比喩を題材とし, BERT の Attention と人間の視線情報から得られた Attention の対応関係を分析することで, 人間の文法的な処理と意味的な処理が BERT のどのような層と対応しているのかを明らかにする.

2)さらに, より人間らしい振る舞いをする大規模言語モデル作成を目指し, BERT における中間の層の Attention の処理に着目し, 人間が採用する情報選択ヒューリスティックの違いを明らかにする.

2.実験概要

2.1 比喩表現データの収集

本研究で扱った比喩表現データは, 比喩表現辞典[2]から収集した. この辞典は, 200 人の作家の 400 編の文学作品から採集した 7,000 件の比喩表現を収録している. また, 青空文庫[3]の原著テキストを用いて, 辞典の文章を 100 文字程度に拡張した. さらに, 各文章に対して, 適切でない 4 つの被喩辞の候補を生成し, その被喩辞を含めた 5 つの文章をデータ化した.

2.2 問題の概要

2.1 節で作成したデータセットの原文の被喩辞部分を 5 つの被喩辞候補（以下, 選択肢）から予測する問題を解かせた. BERT への実験では, 原文の被喩辞部分を空欄にした文章と選択肢を入力し, 空欄に当てはまるものとして最も確率が高い選択肢を出力させた. 同様に, 人間への実験では, 空欄に当てはまるものとして最も適切であると考えた選択肢を回答してもらった.

2.3 BERT の Attention データの収集

2.2 節の問題について, 被喩辞部分を［MASK］に置き換えた原文と選択肢を BERT の学習済みモデル cl-tohoku/bert-base-japanese-whole-word-masking[4] に入力し, 各選択肢が当てはまる確率を出力させた. 東北大学乾研究室で公開されたこのモデルは, 2019.9.1 時点の日本語版 Wikipedia で学習され, 約 1,700 万文で構成されている. 本研究では, 12 層で構成されたモデルアーキテクチャを各層 4 つで構成されるカテゴリーに分類し, 順に浅い層, 中間の層, 深い層と定義した.

また, BERT に搭載されている各単語に注目して予測を行う「自己注意機構 (Self-Attention Mechanism)」という機能を用いて, 2.2 節で予測された最も確率が高い選択肢の各 Token への Attention を各層で取得し, 割合を算出した.

2.4 人間の視線 Attention データの収集

実験では, 被験者に 2.2 節で説明した問題を計 5 問, 24 名に実施した. また, 実験後にアンケートとして, 回答の際に「意味的に注目した」単語を選択させた.

人間の思考時の視線情報は, 視線計測装置 Eye Tracking（Tobii Pro X3-120）を用いて測定した. 被喩辞部分を空欄にした原文を[図 1]のように一文字ごとに四角の枠に等間隔に区切り, 各文字の枠に視線が入

A Comparison of General Purpose Natural Language Processing Models and Human Attention in the Generation of humane Metaphors – Using Gaze Information –

† Wataru Nakamura, School of Management, Tokyo University of Science
‡ Yusuke Ogihara, School of Management, Tokyo University of Science

った回数を数え上げた．これを 2.3 節の BERT の際に分割した Token に合わせて文章を分割し，各 Token に対応する文字を合算し，それに対する視線の回数（以下，視線 Attention と定義する）及びその割合を算出した．

図 1　提示した問題の一例

3.Attention と人間の意思決定の対応関係について

BERT が，どの層で言語構造や品詞に基づいた文の構造を解析し，適切な文法規則を理解するか検証するため，文法的な処理について人間との比較を行った．分析では，文法構造の理解の役割を果たす品詞の特性から，BERT の各層の Attention と視線 Attention を品詞分布で表現した．品詞分布は，MeCab[5]を用いて各 Token に品詞を付与し，それらの品詞の合計値から割合を取得した．そして，被験者の視線 Attention ごとに，各層との品詞分布の KL ダイバージェンス（KLD）を算出した．これを標準化した後，問題ごとに平均を取得し，層ごとの推移を比較した．結果として，KLD と層の深さの間の相関係数は 0.561 であり，正の相関関係を観測した[図 2]．これは，浅い層が文法的な処理に近く，深い層になるほどその処理は，文法的な処理とは遠くなると解釈できる．

また，人間が意味的であると捉えた単語や表現に対して，BERT が認識し，どの程度の Attention の重みを配分するのかを検証した．分析では，アンケートで3人以上の被験者が意味的な解釈上重要と回答した Token を抽出し，それらへ BERT が付与した Attention の重みの推移を検証した．検証の結果，層が深くなるほど，意味的な Token への Attention の重みが増えていることが明らかになった(層の深さと重みの相関係数は 0.634）[図 3]．これは先述の文法的な処理とは反対に，意味的な処理が BERT の深い層で行われていることを示唆している．

Attention の分析においても，先行研究と同様，浅い層ほど文法的な処理が機能し，深い層ほど意味的な処理が機能していることを確認した．ただし，BERT の浅い層から深い層へと相関関係が単調に減少・増加しているのではなく，中間の層では上下していることから，中間の層の処理には依然としてブラックボックスの性質が存在する[1]．

4.BERT のヒューリスティックの再現性検証

BERT の Attention における中間の層の重み配分は，認知科学的な解釈の難しい点の一つであり，先行研究においても明確な結果が得られていない．そこでここでは，人間の情報選択ヒューリスティックとの対応関係を検討した．分析にあたり，2.4 節の結果から被験者を成績上位，中間，下位群に分け，各群から 5 名ずつ抽出した．そして，各成績群の視線配分（Gaze）のエントロピーを，問題ごとに算出した．結果として，問題の順序とエントロピーの間には，上位群は-0.719，中間群は-0.193，下位群で 0.031 の相関係数を観測した．上位群においては，問題の反復によって，視線配分のエントロピーが著しく減少する傾向を確認した．上位層においては，問題の反復によって，視線配分のエントロピーが著しく減少する傾向を確認した．これは問題のパターンや解決方法の学習により，視線の動きが規則的になり，その分布が安定することを示しており，人間が認知負荷を軽減するためのヒューリスティクスを採用しているものと解釈できる．

その一方で，BERT の Attention の重みは，Token の種類や課題によって，中間の層で増加する場合も多々あり，安定しているとは言い難く，浅い層から深い層にかけて情報を集約していくといった人間のヒューリスティクスとは異なる原理を採用していると推察される．今後，言語モデルに人間らしさを導入する際には，この点が一つのベンチマークになるものと思われる．

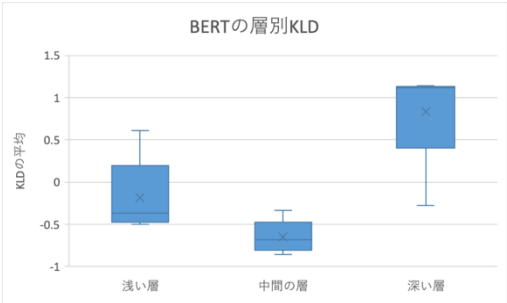

図 2　BERT の層別 KLD

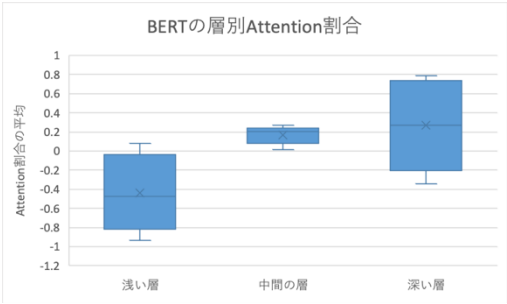

図 3　BERT の層別 Attention 割合

[主要参考文献]

[1] 船井 正太郎，近添 淳一，持橋 大地，浅原 正幸，松井 鉄平，鹿野 豊，川島 寛乃，磯 暁：人間の脳と人工知能における短歌の鑑賞に関する神経活動の比較，言語処理学会第 29 回年次大会 B5-2, 2023.
[2]中村明，「比喩表現辞典」，角川学芸出版, 1995
[3]https://www.aozora.gr.jp/，青空文庫
[4]東北大学 自然言語処理研究グループ：cl-tohoku/bert-base-japanese-whole-word-masking,https://huggingface.co/cl-tohoku/bert-base-japanese-whole-word-masking, 2022
[5]mecab-ipadic-neologd,https://github.com/neologd/mecab-ipadic-neologd