

BERT を用いた語連想タスクにおけるカテゴリ指定効果の分析

Analysis of effect of category designation for word association task with BERT

富島 諒† 相馬 佑哉† 堀内 靖雄† 黒岩 眞吾†
Ryo Tomishima Yuya Soma Yasuo Horiuchi Shingo Kuroiwa

1. はじめに

我々は、人が複数の刺激語から一つの連想語を想起するタスクを計算機で模擬する検討を行なっている。これまで、BERT[1, 2]を用いて「刺激語、…、刺激語から連想する言葉はMASKです。」と入力することでMASKとして連想語を出力させる検討を行ってきた[3]。この際、言葉の代わりに国名等の正解語のカテゴリを指定する語を用いること（以下、カテゴリ指定有）で正答率が向上・低下するものがあつた。そこで本稿では、カテゴリ指定の効果进行分析するため、カテゴリ指定有／無での出力語の違いの分析や、MASKをqueryとするattention weightの比較・可視化を行った。また、カテゴリ指定有で正答率が低下してしまう指定語に対して、指定語を類語に置き換えて実験し、正答率を比較した。さらに、BERTが刺激語と正解語の関係を捉えることができているのかを分析するために「刺激語、…、刺激語から連想するMASKは連想語です。」と入力し、刺激語と連想語からカテゴリ指定語を予測できるかも調査した。

2. カテゴリ指定による効果

本稿では、BERTのMasked Language Modelを用いて5つの刺激語から1つの正解語を予測するタスクを行う。実験で用いる語連想課題として『CD版そのまま使える失語症教材 2』[4]中の『名詞の想起』より、5つの語から特定の語（以下、正解語と呼ぶ）を想起させる『まとめる語想起』を利用した。実験では、『まとめる語想起』132題のうち、20代男子大学生3人の解答が正解語と一致し、かつ、カテゴリが単語で表せる80題を使用した。表1に実験で使用した文の例を示す。カテゴリ指定無の場合は表1の1、3行目に示すように言葉を使用し、カテゴリ指定有の場合は2、4行目に示すように[4]で使用されていた正解語が所属するカテゴリを指定する語、例えば正解語が白ならば所属すると考えられるカテゴリの色を使用する。

評価指標は、hits@1とhits@5を使用する（以下スコアと呼ぶ）。hits@kとは全入力文のうち、k番以内に正解語を出力した文の割合を表す。

2.1 カテゴリ指定有／無の語連想実験結果

表2にカテゴリ指定有／無の場合の予測結果のスコアを示す。赤字はカテゴリ指定有の場合でスコアが低下したものを表す。カテゴリ指定有の場合、色や都道府県等では大きくスコアが向上する一方で、仲間ではスコアが低下した。以下、結果の詳細な分析として、3節でカテゴリ指定有の場合にスコアが向上したカテゴリ指定語の分析を行い、4節でカテゴリ指定有の場合にスコアが低下したカテゴリ指定語の分析を行う。5節では刺激語と正解語からカテゴリ指定語を答えさせる実験を行う。

表1 カテゴリ指定有／無での入力文の例

無	雪、ウサギ、塩、砂糖、紙から連想する言葉は「MASK」です。
有	雪、ウサギ、塩、砂糖、紙から連想する色は「MASK」です。
無	車、電車、飛行機、バス、船から連想する言葉は「MASK」です。
有	車、電車、飛行機、バス、船から連想する仲間「MASK」です。

表2 カテゴリ指定有／無によるスコア

カテゴリ (数)	hits@1(hits@5)	
	指定無	指定有
全体 (80)	0.44 (0.74)	0.50 (0.70)
仲間 (38)	0.61 (0.84)	0.37 (0.47)
都道府県 (8)	0.38 (0.75)	0.63 (0.88)
色 (8)	0.00 (0.13)	0.75 (1.00)
部分 (6)	0.33 (0.83)	0.33 (0.83)
国 (4)	0.25 (1.00)	0.75 (1.00)
場所 (4)	0.50 (0.75)	0.75 (1.00)
スポーツ (4)	1.00 (1.00)	1.00 (1.00)
季節 (4)	0.00 (0.50)	0.50 (0.75)
行事 (3)	0.00 (0.67)	0.33 (0.83)
メニュー (1)	0.00 (0.00)	0.00 (1.00)

3. スコアが向上したカテゴリの分析

表3にカテゴリ指定有でスコアが向上した文の出力語を示す。表中「文」は入力文、「有」「無」はカテゴリ指定有／無を各々示す。また、「有」「無」の行の赤字は正解語を表す。カテゴリ指定無の場合、正解語が白となる文では刺激語の1つである雪に関連する語が多く出力されている。このように、ある1つの刺激語に関連する語が出力されやすい傾向が全般的に見られた。一方でカテゴリ指定有の場合は、指定したカテゴリに所属する語（表3の例では色や季節）が出力されやすい傾向にあった。

次に、attentionに着目した分析を行う。図1に、MASKトークンをqueryとした際にBERTの最終層から出力された12個のattention headによるattention weightを平均して可視化した様子を示す（[SEP]トークンのattention weightの値が大きく、その他のトークンのattention weightの可視化の妨げとなるため、[SEP]トークンを除去した後attention weightを正規化して可視化している）。文の右側に上位5つの出力語を表示しており、例文の正解語は白である。カテゴリ指定有にすることでカテゴリ指定語への注意が大きくなっている。また、どちらの場合でも1~2語の刺激語への注意が大きく、特に先頭の語への注意が大きい。

そこで、刺激語提示の順番の影響を分析するために図1の1-3に1-2の刺激語の語順を入れ替えて実験した結果を示す。語順を入れ替えることで1-2に比べ、刺激語への注意がより均等になったが、正解語を出力する順位は低下した。これより、BERTにおいては語順の影響が大きく、刺激語を適切な順に提示すべきであることが示唆された。

†千葉大学 Chiba University

>>>> 0.00 0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40 0.45 0.50 0.55 0.60 0.65 0.70 0.75 0.80 0.85 0.90 0.95 1.00 <<<<<
 1-1 指定無: [CLS] 雪、ウサギ、塩、砂糖、紙から連想する言葉は「[MASK]」です [水、雨、風、水、太陽]
 1-2 指定有: [CLS] 雪、ウサギ、塩、砂糖、紙から連想する色は「[MASK]」です [白、赤、青、黄色、黒]
 1-3 指定有: [CLS] 雪、ウサギ、砂糖から連想する言葉は「[MASK]」です [水、お菓子、太陽、花、氷]

図 1 カテゴリー指定有でスコア向上した文において MASK を query として可視化した attention

>>>> 0.00 0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40 0.45 0.50 0.55 0.60 0.65 0.70 0.75 0.80 0.85 0.90 0.95 1.00 <<<<<
 2-1 指定無: [CLS] バレーボール、水泳、マラソン、体操、サッカーから連想する言葉は「[MASK]」です [スポーツ、体育、夢、ボール、アスリート]
 2-2 指定有: [CLS] バレーボール、水泳、マラソン、体操、サッカーは「[MASK]」の仲間です [私、僕、海、あなた、わたし]
 3-1 指定無: [CLS] 車、電車、飛行機、バス、船から連想する言葉は「[MASK]」です [乗り物、運、タクシー、汽車、海]
 3-2 指定有: [CLS] 車、電車、飛行機、バス、船は「[MASK]」の仲間です [僕、私、わたし、旅行、地球]

図 2 カテゴリー指定有でスコア低下した文において MASK を query として可視化した attention

表 3 カテゴリー指定でスコアが向上した例

	文	
1	無	雪、ウサギ、塩、砂糖、紙から連想する色は「[MASK]」です。
	有	氷、雨、風、水、太陽、花、アイス、冬、星、夏…
2	無	お花見、入学式、新学期、卒業式、チューリップから連想する季節は「[MASK]」です。
	有	花、桜、夢、春、ゆず、クリスマス、チュー、…
3	無	富士山、浜名湖、茶畑、みかん、うなぎから連想する都道府県は「[MASK]」です。
	有	山、水、富士、ひまわり、海、花、風、太陽、…

表 4 カテゴリー指定有でスコアが低下した例

	文	
4	無	バレーボール、水泳、マラソン、体操、サッカーは「[MASK]」の仲間です。
	有	スポーツ、体育、夢、ボール、アスリート、…
5	無	車、電車、飛行機、バス、船は「[MASK]」の仲間です。
	有	乗り物、運、タクシー、汽車、海、鉄道、夢、…

4. スコアが低下したカテゴリーの分析

表 4 にカテゴリー指定有でスコアが低下したカテゴリー指定語 **仲間** の文の出力語を示す。カテゴリー指定無の場合は 5 つの刺激語に関連する語が出力されやすい傾向があるのに対し、カテゴリー指定語 **仲間** を指定した場合は「僕」や「私」等の人に関わる語が出力されやすい傾向がある。これは **仲間** が、「人々の間柄」という意味で用いられる文が BERT[2] を学習した Wikipedia に多かったと考えられる。

次に、図 2 に attention weight を可視化した様子を示す。カテゴリー指定無の場合は刺激語への注意が大きいのに対し、カテゴリー指定語として **仲間** を用いた場合は 1、2 個の刺激語と **仲間** への注意が大きくなっている。これらのことからカテゴリー指定語として **仲間** を用いることで、人に関連する語が出力されやすい傾向にあり、**仲間** というカテゴリー指定語の選択が適切でなかったと考えられる。

そこで、**仲間** に代わり得るカテゴリー指定語（「グループ」、「一部」等）を使った実験を試みた、表 5 に結果を示す。表より **仲間** よりも **種類** や **範疇** がカテゴリー指定語として適切であると考えられる。なお、**範疇** とした場合も、**範疇** への attention は **仲間** と同程度であった。

表 5 仲間の代替となるカテゴリー指定語でのスコア

指定語	hits@1	hits@5
仲間	0.37	0.47
グループ	0.39	0.53
一部	0.39	0.66
カテゴリー	0.61	0.89
分類	0.68	0.92
種類	0.76	0.89
範疇	0.76	0.92

5. カテゴリーを答えさせる実験

BERT が刺激語と正解語の関係を正しく捉えることができているのかを分析するため、カテゴリー指定語を MASK として出力させる実験を行った。その結果 hits@1 は 0.03、hits@5 は 0.12 となった。これより、カテゴリー推定語の推定は BERT にとってより困難なタスクであることが示唆された。なお、正解した例としては「ピラミッド、クフ王、ナイル川、砂漠、ラクダから連想する「[MASK]」はエジプトです。」(MASK=国)等があった。

6. おわりに

本稿では、BERT に 5 つの刺激語から 1 つ正解語を予測させるタスクにおいて、カテゴリーを指定する語を用いることでほとんどの場合 hits@k が向上することを確認した。さらに、カテゴリー指定有/無での出力語および attention の分析を行った。その結果、BERT は全ての刺激語を同等に注視せず、刺激語のうちのいくつかをより大きく注視して予測していた。これは人間の連想方法に近いと言えるかもしれない。また、カテゴリー指定語を出力させる実験を行なった結果、その精度は極端に低く BERT にとっては困難なタスクであることが示唆された。今後は知識グラフを取り入れた語連想手法を検討する。

謝辞

本研究を進めるに当たり、乾・鈴木研究室の訓練 済み日本語 BERT モデルをお借りしました。モデルを公開してくださったことに厚く御礼を申し上げ、感謝の意を表します。本研究は JSPS 科研費 JP20K11860、JP21K02052 の助成を受けたものです。

参考文献

- [1] Jacob Devlin, et al. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," NAACL, 2019.
- [2] 東北大学 乾・鈴木研究室. "Pretrained Japanese BERT models," <https://github.com/cl-tohoku/bert-japanese>, 引用日期: 2022/06/07.
- [3] 相馬佑哉, 他, "Masked Language Model を用いた語群から語の連想の検討," 言語処理学会第 28 回年次大会, 2022.
- [4] 鈴木勉, 宇野園子監修, CD 版そのまま使える 失語症教材 2, エスコアール, 2022 出版予定.