

深層学習を利用したクイズの選択肢生成手法の提案

松田意仁¹ 赤嶺有平² 根路銘もえ子³

概要: 本研究では空欄に入る単語を推定する穴埋めモデルの出力を利用し、選択式クイズにおける選択肢の生成を行う。選択肢の中に類語が出現することを避けるため、それぞれの単語における分散表現の \cos 類似度やユークリッド距離の閾値を決定し、それをもとに類語判定を行なった。また、穴埋めモデルの出力から選択肢生成に利用する単語数を増やすことで選択式クイズとして利用可能なものが得られやすくなる結果となった。

キーワード: 深層学習, BERT, クイズ

A Method to make quiz choices using deep learning

OKUTO MATSUDA^{†1} YUHEI AKAMINE^{†2}
MOEKO NEROME^{†3}

Abstract: In this study, we propose a new method to automatically make quiz choices using outputs of Masked Language Model that predicts the answer for cloze test. Our method uses synonym identification to remove inappropriate choices based on cosine similarity and Euclid distance between each word-embeddings. In addition, we have got the result that more applicable quiz choices are generated by increasing words taken from Masked Language Model.

Keywords: Deep Learning, BERT, Quiz

1. はじめに

歴史博物館では様々な形でそれらの説明パネルが用意されており、来館者は展示に対する知識を獲得することができる。しかし説明文の全てには目を通さず、重要な箇所を読み飛ばしてしまうことも予想される。そこで展示物に関する選択式のクイズを用意することで、来館者を説明パネル内の重要な部分に誘導できると考えられるが、全ての展示物に対しクイズの問題文や答えの選択肢を人手で準備するにはコスト等の要因により現実的ではない。

The Stanford Question Answering Dataset(SQuAd)[1] のような質疑応答データセットをもとに質問生成モデルを学習させることは可能であり[2], クイズの問題文についてはこれらを利用することで自動生成できると考えられる。

本研究ではクイズの選択肢を自動生成することを目指し、事前に用意した文章に対してクイズとして適した選択肢が生成可能か実験を行う。選択肢の生成には東北大学の乾研究室が公開している BERT(日本語 BERT)から得られる分散表現や、日本語 BERT をもとに学習された空欄に入るトークンを推定する Masked Language Model(穴埋めモデル)の出力を用いる。

2. 提案手法

クイズの選択肢として最低限必要な条件は、日本語の文章として違和感がないことである。Masked Language Modelingを学習した穴埋めモデルはWikipedia等の実際の文章から学習データが作られており、周辺単語と共に用いられる単語が出力されやすくなると考えられ、この条件を満たしやすい。しかし、似た意味の単語が穴埋めの候補として選ばれることがあり、このような意味的な重複はクイズの選択肢として不適切である。そこで類語が複数選ばれる可能性を踏まえ、それらを候補から除外した上で適切な選択肢生成を行う。

2.1 類語判定

2つの単語が類語かどうか判定するために、本実験では日本語BERTから得られる分散表現の \cos 類似度とユークリッド距離を利用する。 \cos 類似度が高い場合、またはユークリッド距離が小さい場合は類語である可能性が高くなると考えられ、事前に類語どうしの値を確認することで適切な閾値を設定し、これにより類語かどうか判定することが可能であると考えられる。

2.2 選択肢生成

本研究では以下の手順を実行し、選択肢を生成する。

- (1) 文章から出題したい語句を決め、“選択肢1”として選択肢に追加する。

1 琉球大学理工学研究科情報工学専攻
Ryukyu University, Graduate School of Engineering and Science,
Information Engineering

2 琉球大学工学部
University of the Ryukyus, Faculty of Engineering

3 沖縄国際大学経済学部
Okinawa International University, Department of Economics

- (2) (1)で決めた語句を[MASK]トークンに置き換え、穴埋めモデルへの入力とする。穴埋めモデルは[MASK]に入る単語を出力し、得られる上位N単語を選択肢候補とする。
- (3) 選択肢候補の上位の単語から順に選択肢へ追加する。ただし、既に追加されている選択肢と類語である可能性が高い場合は選択肢候補から除外する。類語判定は単語の分散表現のcos類似度とユークリッド距離をもとに行う。
- (4) 必要な数の選択肢が揃うまで、または選択肢候補がなくなるまで(3)を続ける。

3. 実験

本実験では選択式クイズとして一般的に扱われている4択クイズの選択肢生成を行い、提案手法で得られる選択肢が適切かどうかを確認する。

3.1 類語判定の実験と結果

穴埋めモデルから得られる選択肢どうしの類語判定を行うため、日本語BERTから得られる単語の分散表現のcos類似度、ユークリッド距離の閾値を決定する。

選択肢生成の実験で用いる文には出現しない似た意味の単語を用意し、表1にそれぞれの値を確認した例を示す。ただし(ドイツ, イタリア)の組は実験で用いる文に出現するが、類語でなくても閾値を超える例があることを示すため記載する。

表1の結果から、3.2で行う選択肢生成の実験ではcos類似度の閾値を0.65、ユークリッド距離の閾値を16と設定した。

	cos類似度	ユークリッド距離
(みかん, オレンジ)	0.78	13.01
(法律, ルール)	0.66	16.60
(洋服, 服装)	0.40	21.75
(食器, 皿)	0.54	19.25
(机, テーブル)	0.74	13.38
(道具, 物)	0.67	15.53
(靴, スニーカー)	0.41	21.51
(ドイツ, イタリア)	0.75	15.42

表1. cos類似度とユークリッド距離の例

3.2 類語判定による選択肢生成

類語判定を行った場合と行わなかった場合でそれぞれ選択肢生成を行い、その結果を比較する。クイズの質問文は「空欄に入る言葉はどれ?」と想定し、生成された4つの選択肢がいずれも日本語として違和感なく空欄に入り、選択肢どうしで意味的な重複のないものを4択クイズの選択肢として適切である判断した。実験では、沖縄と関係のあ

る“首里城”や“エイサー”等についての文章を20文用意し、筆者らが選んだクイズとして出題したい箇所を含む文を穴埋めモデルへの入力として選択肢生成を行い、表2の結果が得られた。また、穴埋めモデルから得られる単語のうち上位12単語をもとに実験を行った。

生成された選択肢が適切かどうかを筆者らで判断した結果を表2に示す。また、生成された選択肢の例は表3のようになっている。類語判定の有無によって“政治”と“政策”や“必須”と“不可欠”など意味の近い単語が除外される結果が見られた。一方で、適切な選択肢と考えられる“イタリア”や“イギリス”が省かれる例も見られた。

不適切と判断したものについては、「名護市以北の本島[MASK]では手踊りの伝統エイサーも続けられている。」という入力に対して「北部, 内, 地域, 島」という4択が得られた例があり、“島”という選択肢により“本島島”という日本語の文章においてあまり目にされない表現ができることから不適切と判断した。

	適切	不適切	4択に満たない
類語判定あり	11	1	8
類語判定なし	12	8	0

表2. 用意した20問に対する選択肢生成の結果

穴埋めモデルへの入力	類語判定ありの場合の出力	類語判定なしの場合の出力
(前略) 沖縄では[MASK]の調達が問題となり(後略)	木材, 種子, 苗	木材, 種子, 苗, 資材
(前略) 按司として成り代わり[MASK]などを推し進め(後略)	海外貿易, 政治, 事業, 改革	海外貿易, 政治, 政策, 統治
琉球舞踊同様に三線は[MASK]の楽器とされてきた。	男性, 必須	男性, 必須, 不可欠, 伴奏
(前略) 大衆化する中で, [MASK]などを取り込む例も増えた。	民謡, 女性, 若者, 方言	民謡, 演歌, 女性, 若者
(前略) 本島[MASK]では手踊りの伝統エイサーも(後略)	北部, 内, 地域, 島	北部, 南部, 中部, 東部
国外で踊られた例として, [MASK]やフランスなどがある。	アメリカ合衆国, ドイツ	アメリカ合衆国, ドイツ, イタリア, イギリス

表3. 類語判定による選択肢の変化の例

3.3 穴埋めモデルの上位 N 単語による選択肢生成

3.2では穴埋めモデルの出力のうち上位12単語から類語判定を行なった上で4つの選択肢を選んだ。次に、穴埋めモデルの出力のうち選択肢生成に用いる上位の単語数を、上位4単語から上位40単語まで変化させることで生成される4択がどのように変化するか調べる。穴埋めモデルへの入力には3.2で用いたものと同じであり、類語判定を含む同様の手順で4つの選択肢を生成したところ表4のような結果が得られた。

	適切	不適切	4択に満たない
上位4単語	0	0	20
上位8単語	8	0	12
上位12単語	11	2	7
上位20単語	13	2	5
上位40単語	17	2	1

表4. 用意した20問に対する穴埋めモデルの上位N単語による選択肢生成の結果

表3には上位12単語から選択肢生成を行い4択に満たなかった例が示されているが、それらに対して上位40単語を用いた場合の違いを表5に示す。また、その他にも適切な選択肢が4つ生成された例や、4択には達したが不適切な選択肢が追加された例を表5に示している。

穴埋めモデルへの入力	上位12単語の場合の出力	上位40単語の場合の出力
(前略) 沖縄では[MASK]の調達の問題となり(後略)	木材, 種子, 苗	木材, 種子, 苗, 松
琉球舞踊同様に三線は[MASK]の楽器とされてきた。	男性, 必須	男性, 必須, 専門, 古代
(前略) 沖縄県立女子工芸学校の[MASK]として利用された。	校舎	校舎, 食堂, アトリエ, 予科
国外で踊られた例として, [MASK]やフランスなどがある。	アメリカ合衆国, ドイツ,	アメリカ合衆国, ドイツ, アルゼンチン, 香港
高価な[MASK]皮を張った三線は富裕さの象徴であったとされる。	蛇, 絹, 金	蛇, 絹, 金, 陶

表5. 上位12単語と上位40単語による選択肢生成の比較

4. 考察

3.1の結果から、表1に示しているように(みかん, オレンジ)や(法律, ルール)など近い意味の単語はcos類似度が高く、ユークリッド距離は小さくなりやすいと考えられる。しかし(ドイツ, イタリア)についても同様の結果となっており、表3のように適切な選択肢が損なわれる例も見られた。これは本実験で行った類語判定が実際は類語ではなく、日本語BERTの学習データ内で周辺語句が似ている単語に反応しているからであると予想される。これにより適切な選択肢が損なわれることも想定されるが、類語を除外する点に限ればその効果は期待される。

表4, 5に注目すると、穴埋めモデルの出力を多く利用することで4択に達しやすくなっている。これは選択肢候補が増えたことにより適切な選択肢の母数が増え、類語判定の誤りにより適切な選択肢が除外されるという点を補うことができたからであると考えられる。一方で、類語を除外できずに不適切と判断された例が見られることから類語判定の精度を高める必要があると考えられる。また、日本語としての自然さを測ることで不自然な選択を除外し、それにより選択肢の質を高めることができると考えられる。

5. おわりに

本研究ではMasked Language Modelの出力を利用した4択クイズの選択肢生成を行った。単語の分散表現を比べることによって選択肢どうしの意味的な重複を避けることができ、4択クイズとしての利用可能性が高まった。しかし、選択肢が空欄に入り自然な文となることを適切な選択肢かどうか判断する指標としており、クイズとしての面白さは考慮されていない。また、4択クイズの自動生成を目指す上で出題箇所を自動決定する仕組みを用意することも今後の課題である。

謝辞 本研究の一部は JSPS 科研費 JP19K01142, JP19K01145 の助成によるものである。

参考文献

- [1] Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, Percy Liang : SQuAD: 100,000+ Questions for Machine Comprehension of Text, Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, 2383-2392, 2016
- [2] Weizhen Qi, Yu Yan, Yeyun Gong, Dayiheng Liu, Nan Duan, Jiusheng Chen, Ruofei Zhang, Ming Zhou : ProphetNet: Predicting Future N-gram for Sequence-to-Sequence Pre-training, arXiv:2001.04063v3, 2020