

汎用言語モデルに co-matching model を追加した 多肢選択機械読解に関する一検討

藤田 和成

荒井 秀一

東京都市大学大学院総合理工学研究科

1 はじめに

機械読解とは、機械が人間と同じように、与えられた文章の内容を理解し解釈するまでの一連の処理であり、与えられた文章に対する質問の正解率で理解度を測るのが一般的な評価方法である。

SQuAD[1] のような機械読解では表層的な質問文が多く、質問文に対する解答が本文中に明示されている設問を対象として、本文中の1文から解答を見つけることで解けるような文章理解を必要としない単純なタスクとなっている。最近、このような単純なタスクに対しては人間と同程度の精度が確認されるようになってきた。[2]

機械が人間と同じように、与えられた文章の内容を理解し質問に答えられるようにするには、本文の内容理解が必要な要素の一つである。本文の内容理解が必要な機械読解の実現には、機械読解タスクの中でも複数の文を組み合わせて解答を行う形式の設問への対応が必要と考えられる。

最近では、本文の内容を十分理解しないと解けないような難易度の高いタスクに取り組む研究もさまざま研究され始め、そのようなタスクの一つに DREAM[3] がある。

DREAM は、詳細なマルチターンマルチパーティダイアログの理解に焦点を当てたタスクになって、回答の84%は抽出型の設問ではなく、質問文の85%は一文を超えた推論を必要としている。DREAM タスクは、文章の内容理解が不可欠な難易度の高いタスクである。

このタスクに対して Yang らは、汎用言語モデルの XLNet を用いた手法を提案している [4]。しかし、この手法では、質問文と選択肢を1つに連結して本文とのマッチングを行なうため、質問文を無視し選択肢の単語に影響を受け誤答することがある。

本研究では本文-質問文と本文-選択肢の各々の重要箇所の抽出を独立して行うことが重要と考え、Co-matching model[5] を汎用言語モデルに追加し、質問文と選択肢を明示的に2つのシーケンスに分けて本文とのマッチングする方法を提案し、その有用性を検討する。

2 Co-matching model

多肢選択機械読解タスクに対して、ニューラルネットワークを用いたさまざまな機械読解モデルが提案されている。その多くのモデルで質問文と選択肢を一つのシーケンスとして扱い、本文とのマッチングを行う手法が多い中、Wang ら [5] は本文-質問文-選択肢を明示的に分け、本文-質問文と本文-選択肢でのアテンションを利用する Co-matching model を提案し、RACE タスク [6] において高い精度を実現した。しかし、単語 embedding の表現能力が低いため、単語やセンテンス間の繋がりを十分に表現できないという問題点を有している。すなわち、類義語などの本文内で同じ意味を持つ単語や語句の言い換えに対応できないため、Dream のような文章の内容理解が必要なタスクに対応できない。

3 XLNet

そこで近年では、大規模なテキストコーパスを用いて語の分散表現を生成した汎用言語モデルで研究が行われている。大きな特徴として語句の言い換えや類義語に強く、単語やセンテンス間の繋がりを表現できるため、Dream のような文章の内容理解が必要なタスクにも対応できる。

そのような汎用言語モデルの一つである XLNet[4] は、大規模なテキストコーパスを用いて事前学習を行った後、タスクごとに fine-tuning する汎用言語モデルである。感情分析、質問応答、含意関係認識などの多くの言語処理タスクで高い性能を発揮している。しかし、2文間の関係考慮タスクに用いられていたため、質問文と選択肢を一つのシーケンスとして扱っている。それが原因で質問文を無視して、本文に含まれる単語に影響された選択肢を選択して誤答することがある。

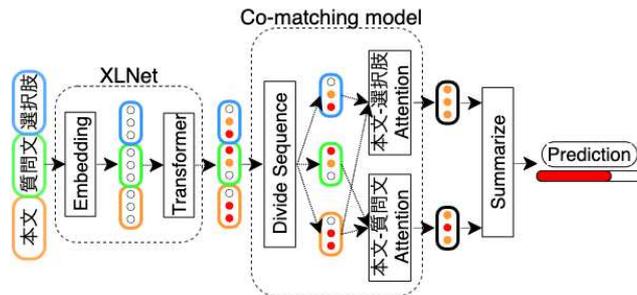


図 1: 提案モデル図

Improving pretrained model with co-matching model for multi-choice reading comprehension.

Kazumasa Fujita, and Shuichi Arai.

Tokyo City University Graduate School

W: How have your two sons been doing at school lately, Andy? M: Terrible! James never starts learning, and Malcolm never stops learning. W: You're joking, of course. I hear that Malcolm is likely to win all the prizes in the exams this year. M: Yes, so his teachers say. He has always been working very hard. He wants to go to Oxford University next year. W: Maybe he'll become a university professor in the end. (後略)
Q1: What is the older son going to do? 1: He wants to win all the prizes in the exam. △ 2: He wants to be a professor in university. 3: He wants to go to university for higher education. ○
Q2: What is the older son going to do? 1: He wants to win all the prizes in the exam. 2: He 'll become a university professor in the end. △ 3: He wants to go to university for higher education. ○

図 2: 正解選択肢の推定例 (従来法: △, 提案法: ○)

4 XLNet と Co-matching model の組み合わせ

本稿では, 図 1 に示すように, 単語やセンテンス間の関係を XLNet で表現し, その出力を本文-質問文, 本文-選択肢を明示的に分離した Co-matching model に接続することで問題点の解決を図る。

5 実験結果と考察

提案手法の有用性を確認するため, DREAM タスクの Test データ 2041 問を用いて, 従来手法の XLNet と正解率を比較した。その比較結果を表 1 に示す。従来手法の XLNet に Co-matching model を追加し, 本文-質問文-選択肢の三つを明示的に分割してマッチングを行うことで, 従来手法の XLNet と比べ 3.8point の精度向上が確認できた。

表 1: 従来手法と提案手法の正解率

モデル名	正解率
XLNet(従来手法)	72.0%
提案手法	75.8%

提案手法で改善した正解選択肢の推定例を図 2 に示す。従来手法の XLNet では質問文 Q1 の考慮度合が本文より低く, 本文中の単語に影響を受け誤答した。そこで, 本文中の単語影響度を確認するため, Q1 の一選択肢を本文の一文を引用したものに変更して実験した。その結果, 従来手法の XLNet では本文の一文に変更した選択肢を正解選択肢と予測し, 本文に含まれる単語に影響されていることがわかった。

一方, 提案手法では, Q1 と Q2 のどちらにおいても本文に含まれる単語に影響された選択肢に惑わされずに, 正解選択肢を選択することができた。よって, 本文-質問文-選択肢を明示的に分けることで質問文を考慮して正解選択肢の予測ができていると考えられる。

本文-質問文-選択肢を三つに明示的に分けても改善できない問題として, 本文内の話者の認識ができていない

M: My school was set up in 1987. W: My school was set up in 1985.
Q1: When was man's school set up? 1: In1987. 2: In 1986. 3: In1985. △○
Q2: When was woman's school set up? 1: In 1987. 2: In 1986. △ 3: In1985. ○

図 3: 話者誤認識に起因する誤答例 (従来法: △, 提案法: ○)

ことが原因で, 誤った選択肢を選ぶことがあった。その一例を図 3 に示す。図 3 の Q2 のみ提案手法では正解選択肢を予測できたが, Q1 に関しては誤った選択肢を選択している。一つの原因として, 本文を一文として扱っているため, 本文の会話内において, どの話者の発話なのか, ある話者の発話範囲がどこまでかを正確に認識できていないのではないかと考えられる。

6 まとめ

本研究では, 質問文を考慮し本文内の単語に影響を受けずに正解選択肢を推定することを目的に, 従来手法の XLNet に Co-matching model を追加する方法を提案した。実験の結果, 従来手法の XLNet に比べ 3.8point の精度が向上し提案手法の基本的な有用性を確認できた。また一例を挙げ本文内の単語に影響されずに正解選択肢を選ぶことも確認できた。

参考文献

- [1] Rajpurkar, P., Zhang, J., Lopyrev, K. and Liang, P.: SQuAD: 100,000+ Questions for Machine Comprehension of Text, in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pp. 2383–2392, Austin, Texas (2016), Association for Computational Linguistics.
- [2] Devlin, J., Chang, M.-W., Lee, K. and Toutanova, K.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pp. 4171–4186, Minneapolis, Minnesota (2019), Association for Computational Linguistics.
- [3] Sun, K., Yu, D., Chen, J., Yu, D., Choi, Y. and Cardie, C.: DREAM: A Challenge Data Set and Models for Dialogue-Based Reading Comprehension, *Transactions of the Association for Computational Linguistics*, Vol. 7, pp. 217–231 (2019).
- [4] Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R. R. and Le, Q. V.: XLNet: Generalized Autoregressive Pretraining for Language Understanding, in Wallach, H., Larochelle, H., Beygelzimer, A., Alché-Buc, d'F., Fox, E. and Garnett, R. eds., *Advances in Neural Information Processing Systems*, Vol. 32, pp. 5753–5763, Curran Associates, Inc. (2019).
- [5] Wang, S., Yu, M., Jiang, J. and Chang, S.: A Co-Matching Model for Multi-choice Reading Comprehension, in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pp. 746–751, Melbourne, Australia (2018), Association for Computational Linguistics.
- [6] Lai, G., Xie, Q., Liu, H., Yang, Y. and Hovy, E.: RACE: Large-scale ReAding Comprehension Dataset From Examinations, in *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pp. 785–794, Copenhagen, Denmark (2017), Association for Computational Linguistics.