

クイズコンペティションの結果分析から見た 日本語質問応答の到達点と課題

有山知希¹ 鈴木潤^{1,2} 鈴木正敏^{3,1} 田中涼太^{4,1} 赤間怜奈^{1,2} 西田京介⁴

¹ 東北大学 ² 理化学研究所

³ 株式会社 Studio Ousia ⁴ 日本電信電話株式会社 NTT 人間情報研究所
tomoki.ariyama.s3@dc.tohoku.ac.jp {jun.suzuki,akama}@tohoku.ac.jp
m.suzuki@ousia.jp {ryota.tanaka,kyosuke.nishida}@ntt.com

掲載号の情報

31 巻 1 号 pp. 47-78.

doi: <https://doi.org/10.5715/jnlpp.31.47>

概要

質問応答は、自然言語処理における重要な研究テーマの一つである。質問応答の研究は、自然言語処理研究の黎明期である 1960 年代から継続的に取り組まれてきた [1]。また、深層学習技術の進展と言語資源の充実により、質問応答研究は世界的に盛り上がりを見せている。特に、SQuAD [2] のような大規模な質問応答データセットや、BERT [3] の登場により、質問応答研究は飛躍的な進展を遂げた。

ただし、質問応答研究の多くは英語で作成されたデータを用いて実施されており、日本語での質問応答の評価はほとんどなされていない。そのため、日本語での質問応答技術がどの程度発展しているのか、その到達点は明らかになっていない。昨今の深層学習技術を質問応答に適用する方法では、言語の違いによる達成度の差異はあまり着目されてこなかったが、扱える言語表現の違い、学習データなどの知識源の質や量の違いなど、言語が異なることによる影響は十分に考慮すべき課題と考えられる。このような背景のもと、日本語での質問応答技術が今後発達していくためには、まず日本語を用いた質問応答技術の現状を明らかにした上で、解決すべき課題を明確にすることが必要である。

そこで本論文では、日本語による質問応答技術の現在の到達点と課題を明らかにし、その上で日本語質問応答システムの今後の改善の方向性を示すことを目的とする。これまで日本語の質問応答技術を評価するための評価データは整備されてこなかつ

たが、本論文では評価データとして、著者らが企画し運営してきた日本語質問応答のコンペティション「AI 王 〜クイズ AI 日本一決定戦〜」¹⁾のために作成したデータセットを用いる。また、評価対象の質問応答システムには、これまでに実施されたコンペティションに提出されたシステムと、汎用の質問応答システムとして利用できる ChatGPT および GPT-4 を用いる。これらのシステムが出力した全解答に対して、それぞれのシステムの特徴と、正解した、または不正解の問題文の中に共通した傾向があるかといった全数チェックを人手で行い、現在の質問応答技術で正答できる、またはできない問題文の特性を分析する。また、システムの特徴に応じた正解の傾向なども調査し、そこから一般化できる知見がないか考察する。以上の分析や考察を踏まえた上で、日本語質問応答システムの改善の方向性を示す。

参考文献

- [1] Bert F. Green, Alice K. Wolf, Carol Chomsky, and Kenneth Laughery. Baseball: An Automatic Question-Answerer. In **Papers Presented at the May 9-11, 1961, Western Joint IRE-AIEE-ACM Computer Conference**, IRE-AIEE-ACM '61 (Western), p. 219–224, New York, NY, USA, 1961. Association for Computing Machinery.
- [2] Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, and Percy Liang. 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, November 2016. Association for Computational Linguistics.
- [3] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 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, June 2019. Association for Computational Linguistics.

1) <https://sites.google.com/view/project-aio/home>