

# 日本語対話機械読解データセットの構築と 対話への応用可能性の検討

## Constructing a Japanese conversational machine comprehension dataset towards improving dialogue systems

芳賀 あかり<sup>1\*</sup> 稲葉 通将<sup>1</sup>  
Akari Haga<sup>1</sup> Michimasa Inaba<sup>1</sup>

<sup>1</sup> 電気通信大学

<sup>1</sup> The University of Electro-Communications

**Abstract:** This paper constructs a Japanese question answering dataset. Machine comprehension systems have achieved remarkable results recently. In addition, machine comprehension in conversations is garnering much attention. However, most of the machine comprehension datasets are in English, and those in Japanese are not known. If the machine comprehension in conversations is applied to the response generation model in the dialogue system, it may be possible to generate appropriate responses after understanding the context. In this paper, we construct a dataset of Japanese machine comprehension in conversations and examine the applicability of the response generation to a dialogue system.

## 1 はじめに

近年、文章と質問が与えられ質問に回答するタスクである機械読解の研究が盛んに行われている。このタスクは対話にも応用されており、対話と質問が与えられ質問に回答する対話機械読解の研究が行われている。しかし、対話機械読解のデータセットは英語のものがほとんどで日本語の対話機械読解のデータセットは知られていない。

本研究では日本語の対話機械読解のデータセットを構築する。具体的には表 1 に示すような対話文に関する質問およびその回答を作成する。また、対話機械読解を対話システムの応答生成に応用すれば、対話文脈を理解した上で適切な応答ができる可能性がある。しかしながら対話機械読解をどのように対話システムの応答生成に応用できるかはあまり検討されていない。本研究では、構築したデータセットを用いて対話機械読解を応答生成に利用した対話システムを提案し、対話システムの応答生成への応用可能性を検討する。

表 1: 対話データの例

| 話者 | 発話                   |
|----|----------------------|
| A  | こんばんは。今日の晩御飯はなんでしたか？ |
| B  | 惣菜店で買った惣菜です。         |
| A  | 惣菜良いですね。なんのお惣菜ですか？   |
| B  | 白身魚のフライと肉味噌炒めでした。    |
| A  | いいですね、魚と肉どちらが好きですか？  |
| 質問 | B はいつ肉味噌炒めを食べましたか？   |
| 回答 | 今日の晩御飯 / 晩御飯         |

## 2 提案手法

### 2.1 データセットの構築

#### 2.1.1 基本設計

日本語対話読解データセット構築のため、クラウドソーシングで収集された雑談対話データ [1] を利用する。対話データの例を表 1 に示す。この対話データから、対話に紐づく質問と回答をクラウドソーシングにより作成する。機械読解のオープンデータセットである SQuAD [2] は、Wikipedia の記事に関する質問と回答をクラウドソーシングによって作成している。SQuAD は「与えられた文章からは答えられない質問」をデー

\*連絡先：電気通信大学  
〒182-8585 東京都調布市調布ヶ丘 1-5-1  
E-mail: h1810750@edu.cc.uec.ac.jp

タセットに含んでおり、このような質問には「回答できない」と回答するモデルを構築している。このような質問に回答できるか判断できるモデルは2.3で説明する質問応答への応用に利用できる可能性があるため、本研究では、「与えられた対話文から答えられる質問」と「与えられた対話文からは答えられない質問」の両方を作成する。

「対話から答えられる質問」の作成では質問に対する回答が対話文の中の連続した文字列で表せるものだけに限定し、回答とセットで作成する。「対話からは答えられない質問」に対する回答は全て「答えられない」となる。

対話機械読解のオープンデータセットである FriendsQA[5] はテレビ番組 Friends の台本をもとに、クラウドソーシングによって質問と回答を作成している。FriendsQA では「what, where, who, why, How, when」のいずれかキーワードを含むように質問を作成しており、キーワードによって回答が変化することを示している。本研究では「何、どこ、誰、なぜ、どのように、いつ」のいずれかのキーワードを含むように質問を作成し、キーワードによる回答の変化や精度の違いを調査する。

また、FriendsQA[5] では一つの質問に対して回答を複数作成し、モデルの学習時において正解に最も短い回答のみを使用した場合、最も長い回答のみを使用した場合、すべての回答を使用した場合に分けて実験を行っている。この実験から、モデルと回答の性質の組み合わせによって精度が変化することが分かっている。これらの理由から、本研究では一つの質問につき回答を複数記載できるものとする。

### 2.1.2 質問と回答の作成

本研究では、クラウドソーシングサイトであるクラウドワークス<sup>1</sup> を用いて対話に関する質問と回答の作成を依頼した。具体的には2人の対話文を読んでもらい、3つの「対話から答えられる質問」とそれぞれの質問に対する回答を作成した。また、2つの「対話文から答えられない質問」も作成した。記入用紙とガイド、記入例をエクセルで作成し、クラウドワーカーにはガイドと記入例を読んでから記入用紙に記入するよう指示した。質問の作成では、キーワードを「いつ、どこ、誰、何、なぜ、どのように」の6つからランダムで1つ指定し、指定されたキーワードを質問文中に含めるよう指示した。

また、対話文中にある質問と同じ意味の質問は認めないこととした。回答は対話文中の連続した文字列を使用することとした。

<sup>1</sup><https://crowdworks.jp/>

回答の作成において、一つの質問に対する回答が複数パターンある可能性がある。例えば、話者Aが「私が好き食べ物はトマトとナスです」と発言していた場合、「Aが好き食べ物は何ですか」という質問に対する回答は「トマト」「ナス」「トマトとナス」のいずれも正解である。

このような場合もできる限り全てのパターンを収集するため、回答が分割できる場合は全体と分割したものをどちらも記載することとした。

まず50対話文の質問と回答の作成を依頼したところ、回答を対話文中から抜き出すのではなく新しく作成してしまう例や、対話文中と同一の質問を記入してしまう例があったため、記入のガイドを修正し再度依頼した。

質問と回答が適切に作成されているかを確認するため、作成した質問と回答において次のような項目を自動的にテストした。

- 質問に指定したキーワードが入っているか
- 回答が対話文中に含まれる連続した文字列になっているか
- 対話文中の質問と類似した質問を作成していないか

これらの条件のいずれかを満たさない場合、データから除外した。

### 2.1.3 構築したデータセット

構築したデータセットの統計情報を表2に、質問の形式別のデータ数を表3に示す。質問数より回答数が少ない場合があるのは、「対話から答えられない質問」に対する回答を除いた数字であるためである。表3から「どこ」「いつ」「なぜ」のキーワードを含む質問に対する回答数が少ないことが分かる。これは「何」「誰」のキーワードを含む質問に対する回答が複数に分けられることが多いためだと考えられる。例えば「何」を含む質問に対する回答は「AとB, A, B」というように「名詞と名詞」のような形の回答が多い。

クラウドソーシングにより作成した質問と回答は、単純なルールに基づく自動テストのみで可不可を判断しているため、作成の誤りを見逃してしまう可能性がある。そこで作成した質問と回答について、いくつかをサンプリングしどのような誤りがあるか人手で調査した。質問作成の誤りでは対話と同じ質問になっている例が見られたが、誤りの数は回答作成より大幅に少なかった。誤った回答の一部を表4に示す。回答作成の誤りでは表4のように回答を分割できる場合も分けていないことが多かった。また、「ええと」のようなフィ

表 2: データセットの概要

|               |      |
|---------------|------|
| 対話数           | 413  |
| 対話から答えられる質問数  | 2904 |
| 対話から答えられない質問数 | 1140 |
| 回答数           | 4071 |

表 3: 質問と回答の内訳

| 質問の形式 | 質問数 | 回答数 |
|-------|-----|-----|
| 何     | 488 | 565 |
| 誰     | 504 | 542 |
| どこ    | 474 | 423 |
| いつ    | 497 | 453 |
| なぜ    | 476 | 433 |
| どのように | 465 | 515 |

ラーを含んだ長い文章になっている例や誤字がある例もあった。これらの誤りは自動的に検出することが難しいため、人手での評価を行うことが今後の課題である。

## 2.2 対話応答生成への応用

対話機械読解の対話システムの応答生成への応用可能性を検討する。先行研究では過去の対話内容に関連した応答を生成できるモデルが研究されている。しかし、既存の対話応答生成モデルは、ユーザが過去に発言した内容からわかるような質問をしてしまう可能性がある。本研究では対話機械読解を用いてこの問題の解決を目指す。

機械読解のデータセットである SQuAD[2] は、「与えられた文章からは答えられない質問」を含んでおり、その場合質問に対する回答は「答えられない」が正解になる。本研究ではこれを対話の応答生成に応用することを考える。例えばユーザが過去に「私が好きな食べ物はリンゴだ」と発言していたとすると、その後システムが「あなたの好きな食べ物は何ですか」と質問してしまうとユーザは同じような発言を強いられることになる。ここで、システムの応答候補から応答に対する回答が「答えられない」となるものを選べば、「与えられた対話からは答えられない質問」をユーザにすることができる。

このように、対話機械読解を対話システムの応答生成に利用できる可能性がある。対話システムの応答生成への応用可能性を確かめるため、作成した対話機械読解モデルを用いて対話システムを構築し人手により評価を行う。

表 4: 誤った回答の例

| 質問      | 誰がロッククライミングに向いていますか？ |
|---------|----------------------|
| 誤った回答 1 | 女性とか柔らかい方            |
| 誤った回答 2 | ええと、女性とか柔らかい方ですかねえ。  |
| 正しい回答   | 女性、柔らかい方、女性とか柔らかい方   |

## 3 まとめ

本研究では、日本語対話機械読解のデータセットの構築と、対話システムへの応用が最終目標であり、その第一段階として対話に関する質問と回答の作成を人手により行なった。現段階では、質問とそれに対応する回答は同一のクラウドワーカーが作成しているためデータセットに偏りがある。この問題を解決するため、今後は作成した質問の言い換えをすることや、質問の作成者とは別のクラウドワーカー複数名で回答を新たに作成することを予定している。また、対話機械読解のタスクを解きデータセットを評価することが今後の課題である。

## 参考文献

- [1] 東中 竜一郎, 稲葉 通将, 水上 雅博 Python でつくる対話システム Python でつくる対話システム, オーム社 (2020)
- [2] Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, Percy Liang: SQuAD: 100,000+ Questions for Machine Comprehension of Text. In Proceedings of the Conference on Empirical Methods in Natural Language Processing, *arXiv preprint arXiv:1606.05250*, (2016)
- [3] Yiming Cui, Zhipeng Chen, Si Wei, Shijin Wang, Ting Liu, and Guoping Hu.: Attention-overAttention Neural Networks for Reading Comprehension. *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, pp.593–602 (2017)
- [4] Adams Wei Yu, David Dohan, Minh-Thang Luong, Rui Zhao, Kai Chen, Mohammad Norouzi, Quoc V.: QANet: Combining Local Convolution with Global Self-Attention for Reading Comprehension. *Proceedings of the 6th International Conference on Learning Representations*, (2018)

- [5] Yang, Zhengzhe, Choi, Jinho D.: FriendsQA: Open-domain question answering on TV show transcripts, *Proceedings of the 20th Annual SIG-dial Meeting on Discourse and Dialogue*, pp.188–197 (2019)
- [6] Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, *arXiv preprint arXiv:1810.04805*, (2018)