

格フレームを用いた質問生成によって深掘りを行う対話システム

Dialogue system for detailed conversation based on question generation using case frame

古川 智雅^{1*} 吉野 幸一郎^{1,2} 須藤 克仁¹ 中村 哲¹
Tomomasa Furukawa¹ Koichiro Yoshino^{1,2} Katsuhito Sudoh¹ Satoshi Nakamura¹

¹ 奈良先端科学技術大学院大学情報科学研究科

¹ Graduated School of Information Science, Nara Institute of Science and Technology

² 科学技術振興機構 さきがけ

² Japan Science and Technology Agency, PRESTO

Abstract: Question generation is one way to communicate interactively with speakers and deepen the dialogue. In this paper, we propose a question generation method that can deepen the dialogue according to the topic or the focus of dialogue. There are two question types depending on their purposes; questions to have a stimulating conversation and questions to acquire information. In this work, we propose a question generation method based on case frame, a statistical frame of predicates and their argument candidates, to build a question generation that can work on both purposes. We constructed the test set annotated with the timing and the content of the question.

1 はじめに

情報検索や雑談のための対話を行うアプリケーションが近年さかんに開発されている。しかしこれらにおいては、一問一答を越えて、より内容に踏み込んだ対話を行うことが難しい。これは、対話の流れを踏まえ、どの部分に焦点を当てて対話における掘り下げを行うかということが考慮されていないためと考えられる [1]。

ここで言う対話における掘り下げとは、話題や焦点を絞り対話を継続する行為とする。掘り下げにはシステムからの質問が有効であることが知られている [2]。本研究では、ドメインに依存しない格フレーム [3] を用いることで、タスクやドメインにかかわらずに対話の内容を深掘りする質問を生成する対話システムを提案する。

システムの質問生成にはユーザの発話を促すためのものと、情報獲得のためのものがある。前者は主に雑談対話システム [4] や傾聴対話システム [5] といった、具体的な目的を持たない非タスク指向型対話システムで用いられている。これに対し、後者は主に情報獲得のための対話システムや、文章理解のための質問生成を行う対話システム [6][7] といった、具体的な目的を持つタスク指向型対話システムで用いられている。提案手法は、これらのいずれの質問生成も行うことができるシステムを構築することを指向しており、情報獲得を視野に入れながら対話を深掘りする対話システムの構築を目指す。

構築した質問生成を評価するため、対話の文脈に応じて質問すべき点、質問すべき内容を付与したテストセットを作成した。本論文では、このテストセットの

作成手順についても述べる。

2 関連研究

質問生成の先行研究では、タスクの有無にかかわらず、主に二つの手法が提案されている。一つはルールによる質問生成 [8][9] で目的に合わせた文法的に正しい質問生成が可能である一方、ドメインごとに質問生成機能を設計する必要がある、コストが高い。これに対し、近年ニューラルネットワークを用いた質問生成 [4][6][7] が検討されている。オープンドメインなコーパスで質問生成を学習するこれらの手法では多様な質問生成が可能になる一方、文法的な誤りや文としての不自然さが現れやすい。また大量の質問応答形式のデータセットが必要である。

これら従来の方式とは異なり、本研究では格フレームを用いた質問生成を行う。格フレームの獲得には大量のテキストが必要であるものの、質問応答に限らない幅広いテキストが利用可能である。また、基本的に用例ベースの質問生成が行われるため、不自然な質問文などが生成されにくく、どのようなドメインに対しても適用可能であるという利点がある。

3 提案手法

本節では 3.1 節で質問生成手法に関して、3.2 節でテストセットの作成方法に関して述べる。本研究では、日本語解析ツールとして形態素解析器 JUMAN++¹ と格・

*連絡先: 古川 智雅, 奈良先端科学技術大学院大学情報科学研究科, furukawa.tomomasa.fo1@is.naist.jp

¹<http://nlp.ist.i.kyoto-u.ac.jp/index.php?JUMAN++>

入力：お土産を買いたいです。

1. 述語に対する意味カテゴリと格のペアを抽出
出力：<抽象物, ノ>, <場所, デ>, <人, ニ>...
2. 意味カテゴリを疑問詞に置き換え、疑問詞と格のペアを取得
出力：<何, ノ>, <どこ, デ>, <誰, ニ>...
3. 疑問詞と格の頻度の割合がしきい値よりも多ければ質問候補を生成

出力：何のお土産を買いたいですか？

どこで買いたいですか？

誰に買いたいですか？

⋮

- ・ノ格以外：疑問詞 + 格 + 述語の繰り返し
- ・ノ格：疑問詞 + 格 + 名詞と格の繰り返し + 述語の繰り返し

図 1: 提案手法

構文解析器 KNP²を用いた。

3.1 格フレームを用いた質問候補の生成

本研究が提案する対話システムでは、発話から抽出された述語と、その述語に対して存在する格フレームをもとに質問生成を行う。具体的な手順を図 1 に示す。提案手法では、まずユーザ発話から述語を抽出し、抽出された述語に対して係る格候補を格フレームから抽出する (ステップ 1)。次に、格要素の性質に応じて疑問詞に置き換える (ステップ 2)。最後に、生成された質問候補が文脈に対して適当であるかを判定する。ここでは、テキスト中の疑問詞と格のペアの頻度に対してしきい値を決め、しきい値よりも頻度が多ければ質問候補を生成する (ステップ 3)。本手法では生成される質問は一つに限らないため、生成された複数の質問を質問候補と呼ぶ。これにより、入力に含まれる任意の述語に対する質問候補を生成することができる。例えば、図 1 のように「お土産を買いたいです。」という入力に対しては、「何のお土産を買いたいですか?」、「どこで買いたいですか?」、「誰に買いたいですか?」の 3 つの質問候補を生成することができる。

3.2 テストセットの作成

提案手法を評価するため、テストセットの作成を行った。対話データとしては、ガイドと旅行者の間で京都観光に関してやり取りされた音声対話を書き起こした、京都観光案内対話コーパス [10] の一部を用いた。このコーパスの一部に対して、3.1 節のステップ 3 においてしきい値を低く設定した提案システムを用いることで、もともとのコーパスに質問候補が加えられたテキストを生成した。これに対し、「ガイドとして旅行者と対話すると仮定して、自分がガイドなら絶対聞く質問・聞いてもおかしくない質問・聞かない質問の三段階にわけてください。また、もし別の質問候補が思いついたら、直前の発話の述語と述語を一致させた質問を書いてください。」というインストラクションのもとでアノテーションした。この結果作成されたテストセットの例を図 2 に示す。

²<http://nlp.ist.i.kyoto-u.ac.jp/index.php?KNP>

T: 適当でいいんですかね。はい。じゃ。

G: はい。

G: はい。わかりました。

G: では、特に何かもうすでに考えられてるコースとか、

G: 訪問してみたい場所とか、

G: なんかありますか。はい。

T: 紅葉が見たいかなど。はい。

0質問: 何を見たいですか？

2質問: どこで見たいですか？

質問: この紅葉が見たいですか？

0質問: 何で見たいですか？

G: 紅葉、そうですね。秋ですもんね。

図 2: テストセットの例 (「G」はガイド、「T」は旅行者、「2」は絶対聞く質問、「1」は聞いてもおかしくない質問、「0」は聞かない質問、太字はアノテーションによって生成された質問をそれぞれ表す。)

4 まとめ

本稿では格フレームを用いた質問生成手法を提案し、テストセットを作成した。今後はこのテストセットを作成して提案システムの評価・分析を行う予定である。

5 謝辞

本研究は JST さきがけ (JPMJPR165B) の支援を受けた。

参考文献

- [1] Koichiro Yoshino, Yu Suzuki, and Satoshi Nakamura. Information navigation system with discovering user interests. In *Proceedings of the 18th Annual SIGDial Meeting on Discourse and Dialogue*, pp. 356–359. Association for Computational Linguistics, 2017.
- [2] 下岡和也, 徳久良子, 吉村貴克, 星野博之, 渡部生聖. 音声対話ロボットのための傾聴システムの開発. *自然言語処理*, Vol. 24, No. 1, pp. 3–47, 2017.
- [3] Daisuke Kawahara and Sadao Kurohashi. Case frame compilation from the web using highperformance computing. In *Proceedings of the 5th International Conference on Language Resources and Evaluation*, pp. 1344–1347, 2006.
- [4] Yansen Wang, Chenyi Liu, Minlie Huang, and Liqiang Nie. Learning to ask questions in open-domain conversational systems with typed decoders. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 2193–2203. Association for Computational Linguistics, 2018.
- [5] Divesh Lala, Pierrick Milhorat, Koji Inoue, Masanari Ishida, Katsuya Takanashi, and Tatsuya Kawahara. Attentive listening system with backchanneling, response generation and flexible turn-taking. In *Proceedings of the 18th Annual SIGDial Meeting on Discourse and Dialogue*, pp. 127–136. Association for Computational Linguistics, 2017.
- [6] Xinya Du, Junru Shao, and Claire Cardie. Learning to ask: Neural question generation for reading comprehension. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1342–1352. Association for Computational Linguistics, 2017.
- [7] Xingdi Yuan, Tong Wang, Caglar Gulcehre, Alessandro Sordani, Philip Bachman, Saizheng Zhang, Sandeep Subramanian, and Adam Trischler. Machine comprehension by text-to-text neural question generation. In *Proceedings of the 2nd Workshop on Representation Learning for NLP*, pp. 15–25. Association for Computational Linguistics, 2017.
- [8] David Lindberg, Fred Popowich, John Nesbit, and Phil Winne. Generating natural language questions to support learning on-line. In *Proceedings of the 11th European Workshop on Natural Language Generation*, pp. 105–114, Sofia, Bulgaria, August 2013. Association for Computational Linguistics.
- [9] Michael Heilman and Noah A. Smith. Good question! statistical ranking for question generation. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pp. 609–617. Association for Computational Linguistics, 2010.
- [10] 翠輝久, 大竹清敬, 堀智織, 柏岡秀紀, 中村哲. 京都観光案内対話コーパスにおける対話行為タグの設計と分析. *情報処理学会研究報告 音声言語情報処理 (SLP)*, Vol. 2009, No. 10, pp. 39–44, Jan 2009.