

チャットボットの会話文を用いた画像生成手法の提案

岩野 真紀† 吉田 雅裕†

中央大学 国際情報学部 国際情報学科†

1. はじめに

入力された文章に対して、応答文を生成するチャットボットの利用が近年拡大してきている[1]。しかし、従来のチャットボットの多くは文章のみを返答するため、その会話内容は単調である。そこで、本稿では応答文だけでなく、応答文に適した画像を出力するチャットボットを制作し、創造性を刺激するコンテンツの生成を試みる。

2. 提案手法の概要

本稿では、自然言語処理技術による応答文生成モデルと、GANによる画像生成モデルを組み合わせたチャットボットを提案する。図1に、提案方式の処理方法を示す。応答文生成モデルには文章や音楽などの時系列データの扱いに特化したモデルであるSeq2Seq (Sequence To Sequence) [2]を採用する。Seq2Seqの学習用コーパスには文献[3][4]の対話コーパスを適用する。また、画像生成モデルには、水彩画の風合いを持つ画像を文章から生成するCLIPDraw [5]を採用する。

提案方式では、ユーザからの日本語の入力文に対してSeq2Seqによる日本語の応答文を生成する。また、CLIPDrawは英語の文章から画像を生成するモデルであるため、日本語の応答文を英語に翻訳してからCLIPDrawに入力し、生成された画像を応答文とともに表示する。

図2に、提案方式の出力結果を示す。提案方式では、ユーザの入力文に対する応答文を表示するとともに、チャットボットの応答文から推定される水彩画風のイラストを表示することで、文字と画像を組み合わせた楽しいコミュニケーションを実現する。

3. 応答待ち時間を短縮するための工夫

提案方式は、チャットボットの応答文を画像生成モデルに入力して画像生成を行うため、ユ

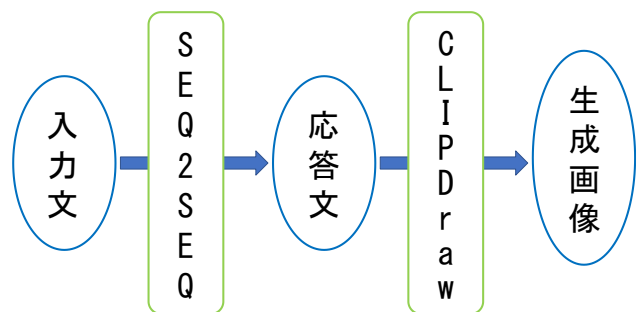


図1 提案方式の処理手順



入力文	海は好きですか？	料理は好きですか？
応答文	海はいいですね	料理は好き

図2 提案方式の出力結果

ーザの応答待ち時間が長くなるという課題がある。例えば、ColaboratoryのGPU環境で提案方式を実行すると、1つの入力文あたり平均390.5秒の応答待ち時間が発生し、円滑な会話を行うことができない。

そこで提案方式では、ユーザとチャットボットの会話内容が一部に偏りやすいという局所性を利用して応答待ち時間の短縮を図る。図3に、応答待ち時間削減のための処理手順を示す。すべての応答文を画像生成モデルに入力するのではなく、頻出する応答文に対応する画像を事前にデータベースに蓄積（キャッシュ）し、頻出する応答文が発生した場合はデータベース内の画像を表示する方式を採用する。

データベースに事前に格納するデータは、過去の応答文をKeyとし、応答文から生成された

A Proposal of a Method for Generating Images Using Response Sentences of Chatbot

† Maki Iwano, † Masahiro Yoshida

† Chuo University, Faculty of Global Informatics

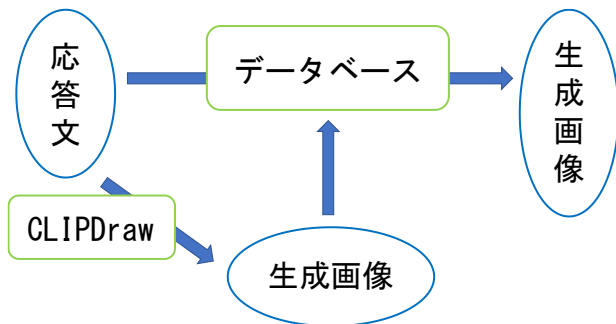


図3 応答待ち時間削減のための処理手順

画像を Value とする。また、データベースの初期状態として、文献[3][4]のコーパス内で出現頻度が高い 100 単語の名詞の中から 2 単語ずつランダムに抽出し、それらを組み合わせた Key によって生成される画像を Value として格納する。

データベースに格納されている画像と、チャットボットの応答文が類似していると判断するための指標として、コサイン類似度を採用する。また、応答文と Key の間で共通している名詞がある場合は、コサイン類似度に 0.4 を加算した値を新たな類似度として用いる。コサイン類似度の算出には、単語の意味をベクトルで表示するための手法である Word2Vec[6]を用いる。コサイン類似度に、共通する単語の重みを加算するのは、類似度の値が文体や構文に左右されるためである。例えば、「海で散歩する」と「散歩をしに海に行きます」という 2 つの文章のコサイン類似度は 0.47 である。しかし、この 2 文を翻訳して生成する画像のどちらにも海や散歩をする人間が含まれ、類似の画像が生成される。したがって、類似度以外にも 2 文章間で共通する名詞という指標が必要となるからである。

本稿では、類似度の閾値を 0.6 に設定し、データベース内に類似度が 0.6 以上の画像がある場合は該当する画像を表示し、類似度が 0.6 未満となる場合はチャットボットの応答文から新たに画像を生成する。その際に生成した画像もデータベースに保存していくことで、平均応答待ち時間の短縮を図ることができる。

4. 評価結果

図 4 に、平均応答待ち時間の評価結果を示す。キャッシュヒット率（類似度の閾値が 0.6 を超える画像が見つかる割合）が 50% のときの平均応答待ち時間は 202.7 秒となった。また、キャッシュヒット率が 100% のときの平均応答待ち時間は 14.9 秒であった。データベースから出力された画像は応答文と共通した単語を反映しているため、応答文に沿った画像となっていた。

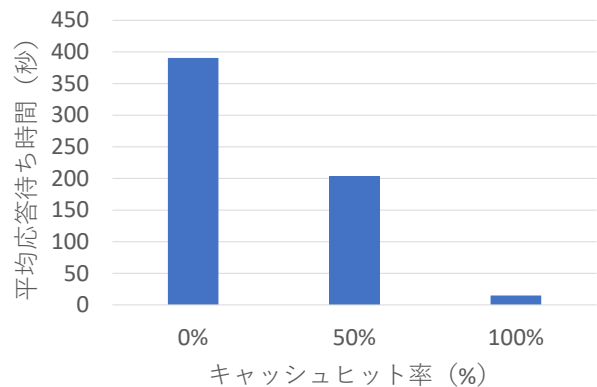


図4 平均応答待ち時間の評価結果

5. 考察

提案手法で採用したデータベースを活用する方式は、応答待ち時間の短縮だけでなく、チャットボットの出力画像の安定性を向上することが考えられる。画像生成モデルが生成する画像は、同じ単語が使用されている場合であっても、単語の順番や文章の長さなどによって生成される画像が変化する。一方、提案手法では、事前に品質の高い生成画像をデータベースに蓄積しておくことで、安定した内容の出力画像を得られると考える。

6. まとめ

本研究ではチャットボットに画像生成モデルを組み合わせることで、応答文に加えて、応答文の内容を反映した画像を出力する手法を提案した。さらに、データベース内に事前に生成した画像を蓄積することで画像生成による応答待ち時間を短縮し、会話の即応性を向上した。

参考文献

- [1] 株式会社矢野経済研究所, “対話型 AI システム市場に関する調査を実施”, https://www.yano.co.jp/press-release/show/press_id/1946 (参照 2023-01-13).
- [2] I. Sutskever et al., “Sequence to Sequence Learning with Neural Networks”, in *Advances in Neural Information Processing Systems* pp. 3104–3112, 2014.
- [3] 東中竜一郎, 船越孝太郎, “Project Next NLP 対話タスクにおける雑談対話データの収集と対話破綻アノテーション”, *人工知能学会 言語・音声理解と対話処理研究会*, vol. 72, pp. 45-50, 2014.
- [4] NTT ドコモ, “雑談対話 API サイト”, https://www.nttdocomo.co.jp/service/developer/smart_phone/analysis/chat/, (参照 2023-01-13)
- [5] K. Frans et al., “Clipdraw: Exploring text-to-drawing synthesis through language-image encoders”, in *arXiv preprint arXiv:2106.14843*, 2021.
- [6] T. Mikolov et al. “Efficient estimation of word representations in vector space”, in *arXiv preprint arXiv:1301.3781*, 2013.