

転移学習を用いた大喜利回答システムの構築

Construction of Ogiri Response System Using Transfer Learning

竹内 海地† 謝 孟春† 中嶋 崇喜† 森 徹†
Kaichi Takeuchi Mengchun Xie Mitsuki Nakashima Toru Mori

1. はじめに

笑いとユーモアは、精神疾患を改善させることが報告されている⁽¹⁾。大喜利はお題に対して面白い回答をするという演芸である。その回答を自動的に作成できれば、利用者はいつでも面白いコミュニケーションを楽しむことができ、利用者の精神衛生の向上も期待できる。

機械学習で大喜利回答システムを構築するためには、大喜利のお題と回答のデータが大量に必要であり、データ収集や学習に非常に時間がかかる。転移学習は、ある問題を効果的かつ、効率的に解くために、別の関連した問題のデータや学習結果を再利用する手法である。そのため、本研究では、少ない大喜利のデータから転移学習によって高精度な大喜利回答システムの構築を試みる。

2. 転移学習による大喜利回答システム

本研究では、図 1 に示した転移学習による大喜利回答システムを構築した。まず、事前学習として、Twitter 上の会話データを大量に集めた大規模な会話コーパスを応答生成モデルの seq2seq によって学習させる。次に、転移学習として、大喜利のお題と回答のデータを集めた大喜利コーパスを、事前学習で得られた重みを初期値として seq2seq によって学習させる。

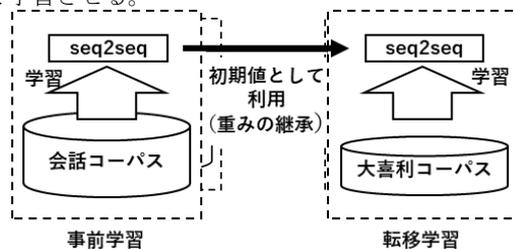


図1 転移学習による大喜利回答システムの構成

事前学習の会話コーパスは、ツイートが発話文、ツイートに付随するリプライを応答文として、発話文と応答文を1つの対とすると、会話コーパスは約150万対の会話データからなる。

転移学習の大喜利コーパスは、誰でも大喜利のお題や回答を投稿できる『匿名大喜利』⁽²⁾というwebサイトからお題と回答のデータを収集し、作成した。お題を発話文、お題に対する回答を応答文とすると、大喜利コーパスは1931対の大喜利データからなる。

3. 実験結果

構築した大喜利回答システムに大喜利のお題を入力し、自動的に生成した応答の一部を表1に示す。応答は概ね自然な回答になっている。

表1の結果について客観的な意見を調査するため、和歌山高専の学生と教員32人にアンケートを行った。質問項目として、回答は面白いかを4段階で評価する。4を「とても面白い」、3を「少し面白い」、2を「あまり面白くない」、1を「つまらない」とする。アンケートの結果を図2に示す。

表1 大喜利回答システムの応答

番号	お題(入力)	システムの応答
1	「牛乳は嗜んで飲むといい」 みたいな事を教えて下さい	五反田住んで敗北はある
2	ゴジラが街を壊すときに心が けている事とは？	すいません
3	ちょっとだけ不安にさせて下 さい	にんにく食べた

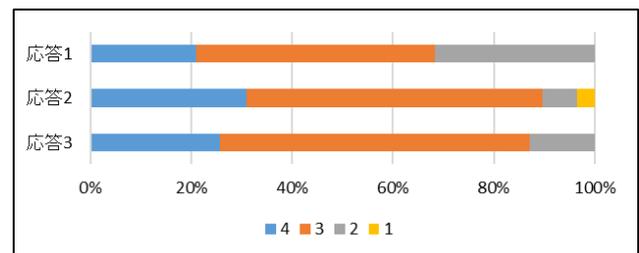


図2 アンケートの結果

図2から、回答1の回答の「とても面白い」が占める割合は21%で3つの中で最も低く、「あまり面白くない」が占める割合は32%で最も高くなっている。また、回答2の回答の「とても面白い」が占める割合は31%で3つの中で最も高く、「あまり面白くない」が占める割合は7%で最も低くなっている。よってお題2に対する「すいません」という回答が最も面白く感じられた。

しかし、回答2では「つまらない」という評価も存在する。これはお題2の「心がけている事」に対して「すいません」という表現が適切でないと判断されたことが考えられる。

4. まとめ

本研究では、転移学習と seq2seq を用いて大喜利回答システムの構築を試みた。その結果、大喜利として面白いと感じられる回答を得られることが確認できた。しかし、日本語として不適切、つまらないと捉えられる場合もある。モデルのパラメータの調整や、データセットの改良などによって、さらに正しい日本語や面白いと感じる大喜利の回答を実現することが今後の課題である。

参考文献

- [1] 森田, 心理的援助への笑いとうもろこしの適用に関する研究の動向と課題, 笑い学研究 25(2018.7)
[2] <https://toogiri.buhoho.net/>