

Transformer 対話モデルによる未来の展開の予測と感情の誘導の検証

進藤優 柴田千尋
法政大学 理工学部

yu.shindo.2h@stu.hosei.ac.jp, chihiro@hosei.ac.jp

概要

近年、過去の文脈を考慮する対話システムの研究が進んでいる。しかし、対話においては未来の展開を考慮することも重要である。そのため、本研究では、未来の展開を考慮し、対話5ターン以内にユーザーがポジティブになる先読み対話システムを提案する。これは、対話システム本体 T_1 とは別で独立した対話システム T_2 を用いて、 T_1 の発話生成時に T_2 がこれに返答をすることで先読みを行う。これをシステムと人間相手に対話させることで、提案システムの有効性、対話の自然性を評価する。

1 はじめに

近年、深層ニューラルネットワークを用いた自然言語生成技術の発展と膨大な対話データの存在に支えられ、対話システムは大きく発展してきた。これにより、日本語においても、りんな [1] などの非タスク型の対話システムが身近な存在となっている。これらの対話システムは、発展した技術に加えて、過去の文脈を考慮することで対話の自然性や一貫性を向上させている。一方で、人間にとってより良い対話システムを構築するためには、過去の対話履歴だけでなく、未来の展開を考慮することも重要である [2]。しかし、未来の展開を考慮する対話システムの具体的な研究はまだ十分に行われているとは言えない。そこで本研究では、未来の展開を考慮した対話システムを提案する。具体的には、未来の展開を考慮し、対話5ターン以内にユーザーがポジティブになる先読み対話システムの作成を提案する。

2 関連研究

2.1 Transformer 対話モデル

本研究では日本語の大規模な対話モデルとして、NTT により公開されている Transformer 対話モデ

ル [3] をベースとして用いる。相手に共感するような対話データ [4] 約 20,000 件からファインチューニングを行っている。1件4発話で構成されるため、計 80,000 発話のデータとなっている。このデータを用い、過去の発話を4発話分まで考慮することで対話の自然性を向上させている。発話の生成は、複数の発話候補を生成し、そこから文章として成り立っている確率が最も高いものを選択して返答する。この流れを図1に示す。

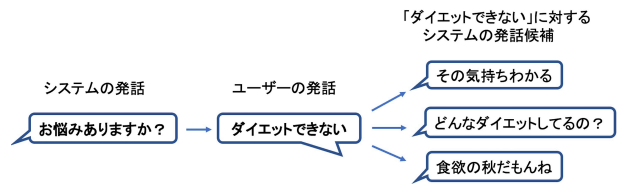


図1 Transformer 対話モデルの対話の流れ

2.2 対話の先読みに関する研究

岸波らは、先読み対話システムの可能性について検証している [2]。この研究では、人間同士で対話を行い、タスクの結果から先読み対話システムの可能性を分析する。この研究で提案されたタスクは、パーティーゲームのひとつである NG ワードゲームを類似したものであり、システムと人間の1対1で対話が行われる。システム側は、1ターン目の発話とターゲット単語が与えられ、対話を続けていく中で5ターン以内に人間からターゲット単語を引き出すことを目指す。この実験により、対話のターン数を N として、 $N=5$ としたとき最も対話の自然性が高いということがわかった。また、設定する先読みゴールの難易度によっては、対話システムで先読みを行うことは可能であるということもわかった。

3 提案手法

本研究では、未来の展開を考慮した対話システムを提案する。先読みのゴールとしては、システ

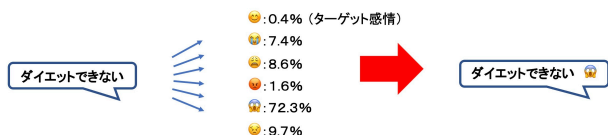


図2 感情分析の流れ

ムに1ターン目の発話及びターゲット感情のペア、(例えば「こんにちは。何かお悩みはありますか?」及び「😊」のペア)を与え、対話を続けていく中でユーザー発話がターゲット感情になることを目指す。未来の展開を考慮し、対話 N ターン以内にユーザーがポジティブになる先読み対話システムを提案する。本論文では、岸波らの研究結果に従い、 $N=5$ とした。

3.1 感情分析

ユーザーの発話がターゲット感情になるように先読みを行うが、そのためにはまず、発話がどのような感情を含意しているのかを推定する必要がある。ターゲット単語を設定して先読みを行なう手法[2]では、数多くある単語の中からピンポイントにターゲット単語を引き出すことは難しい。そこで、本研究では、Twitterの絵文字付きツイートを学習し、文章を6つの絵文字(感情)に分類する感情分析用のBERTモデル[5]を用いる。このモデルは、それぞれの絵文字ごとに別々に予測確率を算出する、いわゆるマルチラベルモデルである。例えば、😊と👍が共に90%以上となることもあり得る。この流れを図2に示す。これにより、あらかじめ設定された6つの絵文字の中からターゲット感情を引き出す先読みタスクを設定することができ、ターゲット単語よりも実現可能であると考えられる。

3.2 先読み過程

既存のTransformer対話モデル[3]をベースとし、それに先読み過程を追加する。具体的には、対話システム本体であるTransformer対話モデル T_1 とは別で独立した対話モデル T_2 を用いる。 T_1 のユーザー発話に対する複数の発話候補それぞれに対して、 T_2 が返答候補を生成する。この返答候補が入力されたユーザー発話のひとつ先の予想ユーザー発話に該当する。この流れを図3に示す。生成された複数の T_2 の返答候補を T_1 の発話候補ごとに区切って評価する。最も良い評価の返答候補の集合を引き出した発話候補が T_1 のユーザー発話への返答として選択される。

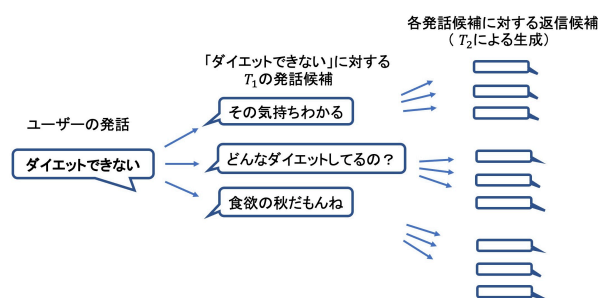


図3 提案システムの対話の流れ

3.3 先読みによる評価値の計算

各 T_1 の発話候補に対して、先読みによって評価値を計算する方法を二つ提案する。一つ目は、各 T_1 の発話候補に対する T_2 返答候補の発話それぞれに対して、感情分析[5]を行い、その感情が何個ターゲット感情「😊」になっているかを比較する方法である。このとき、ターゲット感情になっている数を、その T_1 発話候補の評価値とする。二つ目は、 T_2 返答候補のターゲット感情の予測確率を足した合計値で比較する方法である。この方法では、各 T_2 返答候補において、必ずしもターゲット感情が最も高い予測確率を持っている必要はない。感情分析の過程で求められる6つの絵文字に対する確率の中からターゲット感情「😊」の確率を使用し、返答候補ごとに合計値を求める。その合計値が最も高いものを最も良い評価とする。また、一つ目の方法で最もターゲット感情が多い返答候補が複数出た場合は、その返答候補の中で二つ目のターゲット感情の確率で比較する方法を適用する。

一つ目の方法の評価値は、ターゲット感情がそれ以外の感情と比べて最も強いかどうかの2値の合計値となるため、ターゲット感情のBERTモデルによる分類結果自体に注目することができる。また、二つ目の評価値では、例えば、ラベル付けされた感情は「👍」だがラベル付けされた感情とターゲット感情の確率が僅差などの状況を考慮できるため、細かい情報まで考慮できる。

4 実験

提案手法の評価のため、以下の3つの実験を行う。

- 実験1: Transformer対話モデルと提案システムとの比較
- 実験2: 提案システムの有効性
- 実験3: 提案システムの対話の自然性の評価

4.1 前提条件

今回提案する対話システムの本体 T_1 と先読み部分の対話システム T_2 では、2 節の Transformer 対話モデルと同様に Twitter の対話データを Empathetic-Dialogues [4] でファインチューンした対話モデルを使用する。これにより相手の感情に共感するようになり、対話の自然性も向上する。また、感情分析では町田ら [5] の Twitter から抽出した絵文字付きツイート 360 万件を使って、6 つの感情に分類する手法を使う。

4.2 システムとの対話

2 節の Transformer 対話モデルと提案システムをそれぞれ対話システム（ユーザー側）と対話させる。このときのユーザー役の対話システムも NTT が公開している対話システム [3] を使用する。ユーザー役を対話システムにすることにより、自動で対話させることができるので多くの対話データを取得することが可能になる。このとき取得する対話データは、5 ターン分の対話履歴、最後のユーザー発話の感情ラベル、最後のユーザー発話のターゲット感情の確率とする。

実験 1 では、Transformer 対話モデルと提案システムをそれぞれ、もう一つの別の対話システムと対話させ、取得した対話データから Transformer 対話モデルと提案システムを比較する。

実験 2 では、提案システムのターゲット感情を他の感情に変えることで提案システムの有効性を検証する。実際には、先読みした発話のターゲット感情の確率が最も低いものを選ぶようにする。これにより、ターゲット感情「😊」以外の 5 つの感情に注目するようになる。これをターゲット感情を「😊」にしたときと比較することで、ターゲット感情が「😊」のとき、本研究の目的であるユーザーがポジティブになる先読み対話システムがより実現できているかを確認する。また、実験 1 と同様に対話システムと対話させることで比較をする。

4.3 アンケート調査

実験 1、実験 2 では主にターゲット感情の数やその確率を使って、本研究の目的の達成率を調べている。実験 3 では、ユーザー役を人間が行い、システムにはわからない Transformer 対話モデルと提案システムの違いや対話の自然性について調査する。具

体的には、人間に Transformer 対話モデル、提案システム（ターゲット感情「😊」）、提案システム（ターゲット感情「😊以外」）の 3 つのシステムと対話してもらい各対話終了時にアンケート調査を実践する。アンケートの項目は、「あなたは悩んでいたが今ポジティブになったか」「システム側の発言は自然だったか」について、それぞれ 5 を最高点とした 5 段階評価とする。その結果から Transformer 対話モデルと提案システムの違いや対話の自然性を評価する。

5 実験結果

4 節で述べた実験 1~3 の結果を記述する。実験 1,2 では、比較する対話システムごとに 85 対話のデータを取得した。実験 3 では、5 人の方に協力してもらい、15 対話のデータを取得した。

5.1 Transformer 対話モデルと提案システムとの比較結果

実験 1 の結果を表 1 に示す。このとき、3 節で述べた提案手法の二つの評価方法も比較する。表にある提案手法 1 は、先読みした発話の感情が何個ターゲット感情になっているかで評価をする方法で提案手法 2 は、先読みした発話のターゲット感情の確率で評価する方法である。これを最後のユーザー発話が 85 対話中、何個ターゲット感情になったかとターゲット感情の確率の平均値で比較する。

表 1 Transformer 対話モデル（ベースライン）と提案システムとの比較

| システム | ターゲット感情の数 | ターゲット感情の確率の平均値 |
|----------|-----------|----------------|
| ベースライン | 77 | 83.109 |
| 提案システム 1 | 81 | 86.311 |
| 提案システム 2 | 82 | 88.854 |

表 1 より、Transformer 対話モデルよりも提案手法の方がターゲット感情にさせる数と確率が高いということがわかる。つまり、提案手法の方が本研究のユーザーをポジティブにするという目的を達成していると言える。また、提案システム 1 よりも提案システム 2 の方がターゲット感情にさせる数と確率が高い。そのため、先読み過程では、ターゲット感情の確率で評価した方が良いとわかる。

5.2 提案システムの有効性の結果

実験 2 の結果を表 2 に示す。このとき、ターゲット感情を「😊」とした提案手法を提案システム「😊」

とし、ターゲット感情の確率が最も低いものを選択し、ターゲット感情を「😊」以外とした提案手法を提案システム「😊以外」とする。また、このとき取得したターゲット感情の数は、二つともターゲット感情が「😊」のときのデータである。

表2 提案手法によるターゲット感情への誘導の有効性

| ターゲットとする感情 | 最終的な「😊」の予測数 | 最終的な「😊」の予測確率の平均値 |
|------------|-------------|------------------|
| 😊 | 82 | 88.854 |
| 😊以外 | 69 | 69.671 |

表2より、ターゲット感情を「😊」としたときの方がターゲット感情にさせる数と確率が高いということがわかる。つまり、設定するターゲット感情に影響して結果が変わることが言える。また、これと同時に今回提案した手法が有効であることも確認できる。

5.3 提案システムの対話の自然性の評価

実験3の結果を表3に示す ($n = 15$)。また、表6に提案システムの実際の対話例を示す。このとき、比較するシステムはTransformer対話モデルと実験2で使用したシステムである。

表3 対話の自然性の評価 (5段階評価アンケート平均値)

| システム | ポジティブになったか | システムの発言は自然だったか |
|------------|------------|----------------|
| ベースライン | 2.8 | 3.4 |
| 提案手法 (😊) | 4.2 | 3.8 |
| 提案手法 (😊以外) | 2.4 | 3.2 |

表3より、ターゲット感情を「😊」としたときの提案システムが最も自然な対話でユーザー (人間) をポジティブにさせていることがわかる。最後に、上述の差が有意かどうかをt検定を用いて検証する。具体的には、有意水準を5%としたときの対の標本による平均のt検定をTransformer対話モデルと提案システム「😊」に対して行う。相手がポジティブになったかの結果を表4に対話が自然であったかの結果を表5に示す。

表4 ポジティブになったかの検証結果 (2標本t検定)

| | ベースライン | 提案手法 |
|--------------------|---------|------|
| 平均値 | 2.8 | 4.2 |
| 分散 | 0.7 | 0.2 |
| t | -5.72 | - |
| P($T \leq t$) 両側 | 0.00464 | - |
| t境界値 両側 | 2.78 | - |

表4のP($T \leq t$) 両側の項目からp値は「0.00464」となり、0.05 (有意水準) より小さいので、Transformer

表5 対話が自然性だったかの検証結果 (2標本t検定)

| 結果 | ベースライン | 提案手法 |
|--------------------|--------|------|
| 平均値 | 3.4 | 3.8 |
| 分散 | 0.8 | 0.7 |
| t | -0.667 | - |
| P($T \leq t$) 両側 | 0.541 | - |
| t境界値 両側 | 2.776 | - |

表6 提案システムの対話例

| 提案システム | ユーザー (人間) |
|--|--|
| こんにちは。何かお悩みはありますか？ | ダイエットできなくて悩んでいます。 |
| ダイエットは焦るとリバウンドしやすいですよ。焦らずゆっくりダイエットを始めてみてはどうでしょう。 | でもやる気が出なくて始められるか分からないです。 |
| ダイエットはやる気が出るまでが長いですからね。少し時間を置いたらやる気が出てくるかもしれませんよ。 | 参考にしてみます！でももし時間置いてもやる気が出なかったら、どうしたらいいかな？ |
| うーん、やる気が出るまで他のことをやってみたり、一度、痩せた自分を思い浮かべてみるのはどうでしょう。 | 教えてくれてありがとうございます。参考にしてみます。 |
| 頑張ってください！ダイエット成功したら何かご褒美を考えましょう。 | ありがとうございます！ご褒美のためにも頑張ります。 |

対話モデルと提案システムの平均値には有意差があると言える。つまり、提案システムの方が相手をポジティブにするという結果は信用性があるとわかる。一方で、Transformer対話モデルと提案システムの対話の自然性については、平均値では提案システムのほうが高い評価となっているものの、表5から、その差は有意があるとは言えないことがわかる。

6 結論

本研究では、未来の展開を考慮し、対話Nターン以内にユーザーがポジティブになる先読み手法を提案し、システムとして実装を行い、評価を行った。その結果、提案手法は、システムと人間の両方に対して、先読みを行うことにより、ターゲット感情へと対話を誘導することができることがわかる。さらに、対人の実験結果からも、今回の提案手法がユーザーをポジティブにできていることがわかる。今後の展望としては、感情分析の精度を上げることで、より未来の感情を的確に考慮することが可能か検討する予定である。

7 謝辞

今回の研究でアンケート調査にご協力いただいた法政大学の学生の皆様には、深く感謝申し上げます。本研究の一部は JSPS 科研費 JP18K11449 の助成を受けたものです。

参考文献

- [1] Katsuya Iida, Kazuna Tsuboi, Momo Klyen, Xianchao Wu, Kazushige Ito. りんな：女子高生人工知能, 2016.
- [2] 岸波洋介. 対話システムの先読み能力を分析可能なタスクの検討, 2021.
- [3] Hiroaki Sugiyama, Masahiro Mizukami, Tsunehiro Arimoto, Hiromi Narimatsu, Yuya Chiba, Hideharu Nakajima, and Toyomi Meguro. Empirical analysis of training strategies of transformer-based japanese chat systems, 2021.
- [4] Qintong Li, Piji Li, Zhaochun Ren, Pengjie Ren, and Zhumin Chen. Knowledge bridging for empathetic dialogue generation. 2022.
- [5] 町田秀輔. 絵文字の埋め込み表現を用いたラベル埋め込み注意機構モデル等による文の感情推定, 2019.