

## GPT2 を用いた対話システムの構築

早稲田大学 基幹理工学部 情報通信学科 3 年 1W202254 中村友亮

### 概要

本研究では、GPT2 モデルを用いた対話システムの構築を行った。話者における話しやすさ、自然さ、会話内容の整合性を実現するために、ペルソナの付与を利用して対話実験を行った結果、過去の対話内容を反映した応答の生成に成功したが、一部ペルソナを与えた場合や、入力によっては会話内容に矛盾が生じてしまうことを確認した。

### 1. はじめに

#### 1.1 背景

Pepper, Robi などの雑談ロボットに限らず、人間が対話システムを利用する際、ボットが話しやすく、人間に近い会話を行うことが出来るかどうかを利用者の安心感、使い心地の一因となる。

#### 1.2 目的

対話システムの応答が過去の対話内容を含めて整合性のあるものとなること、また、話者の口調に近いものとなることを目標としてシステムを構築した。

### 2. 関連研究

#### 2.1 ポライトネス理論

人間が会話をする際、「フェイス」という欲求をもち、これを保持しながら会話するように努めるという理論である。フェイスには話し相手から称賛されたいというポジティブな感情による欲求である「ポジティブフェイス」と、他者に干渉されたくないというネガティブな感情による欲求の「ネガティブフェイス」という二種類がある。本研究では、これが話しやすさにつながるという前提のもとで対話システムの構築をおこなう。

#### 2.2 ペルソナを考慮した応答生成

フェイスの概念を対話システムに組み込むために、話者や対話システムの情報、関係などをペルソナとして入力に与え、このペルソナに従うように応答を生成する研究である。また、大規模言語モデルを対話システムとして応用するためにペルソナに矛盾しないいくつかの対話例も入力に加える。

### 3. 方法

#### 3.1 実験モデル

事前学習済みモデルとして rinna の japanese-gpt-1b を使用した。また、実装には Python を用い、Tokenizer に

は transformers から T5Tokenizer を使用した。話者是对話を始める前に年齢、性別、趣味を入力として与え、それをもとにモデルにペルソナと対話例を与える。話し相手の属性が互いに近ければ話しやすさにつながると考え、話者に設定するペルソナと対話システムに設定するペルソナを同一のものにした。対話例は 30 歳以上であるか否か、男女どちらであるかによって 4 種類に分けたもののうち、該当するものを与える。ショット例を図 1 に示す。

A は 40 歳の男で趣味は筋トレである。  
B は 40 歳の男で趣味は筋トレである。  
A と B は友達である

A: こんにちは。元気かい？  
B: うん。まあそこそこに。最近は何してる？  
A: 今は仕事を頑張ってるよ。  
B: いいね。俺も最近仕事が結構忙しくて。  
A: 大変だね。頑張って。  
B: ありがとう。そちらも頑張ってくれ。

図 1. モデルに与えるショット例（A が話者、B がシステム）

また、話者の入力とそれに対するシステムの応答も図 1 の対話例に続けて加える。

### 3.2 実験方法

話者の入力とそれに対するシステムの応答を 10 回繰り返し、それを 30

歳以上であるか否か、男女どちらであるかの 4 つの場合においてペルソナや会話内容を変えて 5 回ずつ行う。評価は応答ごとに筆者の主観によって行う。

一つ目の評価基準は整合性であり、

- ・ 応答が文章として成立しているかつペルソナや過去の会話に矛盾していない場合に 1 点
- ・ そうでなければ 0 点

二つ目の基準は応答の自然さであり、

- ・ 応答が話者の口調にそったものであるならば 1 点
- ・ そうでない場合、例えば話者が敬語でない入力をしているのに対してシステムが敬語で応答するなどしたら 0 点

とし、評価基準ごとに 1 応答当たりの平均値をスコアとする。

## 4. 評価

結果を表 1 に示す。

表 1. 各属性における応答のスコア

属性	整合性	自然さ
30 以上男	0.9	1.0
30 以上女	0.89	0.7
30 未満男	0.72	0.54
30 未満女	0.72	0.86

表 1 から分かるように 30 歳以上の男性の場合において整合性、自然さ共に最も高いスコアを獲得した。各属性におけるスコアの違いの考察として整合性においては、年齢が高い二つの

属性が低い属性よりも高いスコアとなっていることから、30歳以下の属性に用いた対話例に含まれる若者言葉が整合性を満たさない応答を誘発してしまったと考えられる。自然さに関しては、0点だった場合のほとんどが敬語で応答してしまったことが原因であったので、特にスコアの低い30歳以上の女性、30歳未満の男性の場合に与えた対話例が敬語と結びつきやすいものであり、敬語でない応答を生成するには数が不足していたためであると考えられる。また整合性に関して、実験の際は応答が矛盾していた場合に、それをなるべく追求しないような文を次の入力として与えるようにしていたため、もし仮に一般のユーザーが実際に対話システムを用いた場合、整合性のスコアは表1に示した値よりも低くなると考えられる。

## 5. おわりに

実験中、「読書をする」という旨の応答に対して相槌のつもりで入力した「本が好きなんだー」という文に対して話者自身が本が好きであると発言したとシステムが勘違いし、「そうなんだー」と返してしまい整合性で0点を獲得したことがあった。文章としては同じであっても発言に対する立場が異なってしまうこのようなケースは今後の課題の一つであると考えられる。また、一般的に話しやすい存在である友人とは多くの場合敬語を用いずに会話を行うため、今回の実験で

は敬語の応答は不適であると考えたが、人によっては親しい中の友人同士であったとしても敬語で話すような場合も少なくないため、これを考慮した話しやすさの尺度の改善が求められる。

## 6. 参考文献

- [1] 宇佐美まゆみ. ポライトネス理論と対人コミュニケーション研究. 国際交流基金, pp. 7-8, 2002.
- [2] 川本稔己, 山崎天, 佐藤敏紀, 奥村学. 大規模汎用言語モデルによるペルソナを考慮した応答生成. 言語処理学会 第28回年次大会 発表論文集, pp. 1344-1348. 2022.
- [3] 山崎天, 川本稔己, 吉川克正, 佐藤敏紀. 大規模汎用言語モデルを用いた雑談対話システムの対人関係性に基づく発話制御の検討. 言語処理学会 第28回年次大会 発表論文集, pp. 1921-1925. 2022.