

応答履歴に応じたペルソナの更新が対話システムの応答生成へ与える影響の分析

吉田 快¹, 品川政太郎^{1,2}, 須藤 克仁^{1,2}, 中村 哲^{1,2}

1奈良先端科学技術大学院大学, 2理化学研究所革新知能統合研究センター

概要: 応答履歴によってペルソナを更新する対話システム

ペルソナ: I am a student.
 クエリ1: What university do you belong?
 エージェント: I am Tokyo university student.
 ペルソナ: "I am a student.",
 "I am Tokyo university student."
 クエリ2: Where do you live?
 エージェント: I live in Tokyo.

従来のペルソナ対話システム [1] の課題

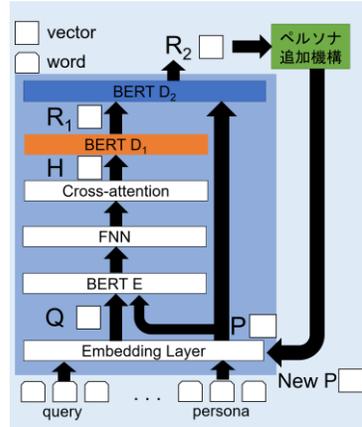
- ペルソナが初めに条件付けされたものから更新できない
- 条件づけられていないペルソナについての質問には、新たなペルソナを含んだ応答を生成する可能性がある

応答履歴によるペルソナ更新の例

本研究のアイデア: ペルソナの更新を考えることで上記の問題を解決できるのでは?

ペルソナの更新における課題

- どのようなペルソナが生成されるか未知 (実験①, ②)
- ペルソナの長さ (実験③), 既存のペルソナとの重複や矛盾 (実験④) による影響が考えられるため, それらの問題を調査する必要がある



応答生成モデル(BoB)とペルソナ追加機構

応答生成モデル(BoB)とペルソナ追加機構

応答生成モデル: Songら[1]のBERT over BERT(BoB)を使用

ペルソナ追加機構

応答履歴から以下の条件を満たすものをペルソナとして抽出し, フィルタリングを適用して新しいペルソナを作成する

1. 名詞, 代名詞, 形容詞のうち少なくともどれか1つを含む
 2. 4から20の単語で構成される
 3. "I"か"my"という単語を含む
- フィルタは重複と矛盾の2種類を想定

実験設定

- 評価データ: Conv AI2 Persona Chat (7,504件の対話に対して応答生成を行って評価)
- 生成された応答候補群からサンプリングする方法として, 応答生成に多様性を持たせるためにtop-pサンプリングを用いた(先行研究ではgreedy search)

○評価指標

distinct-n: 生成された応答の多様性 (情報量) を測る

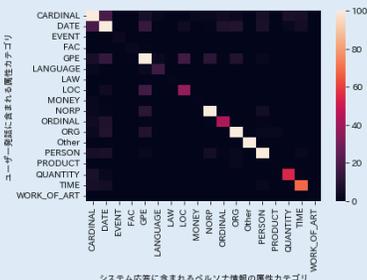
$$distinct - n = \frac{|unique(n - gram)|}{|uni - gram|}$$

C.Score: NLI分類器によって, どの程度ペルソナを含んだ応答生成ができていないかを測る

$$C.Score(r) = \sum_{i=1}^t NLI(r, p_i) \quad NLI(r, q_i) \begin{cases} -1 & r \text{ が } p_i \text{ と矛盾} \\ 0 & r \text{ が } p_i \text{ と中立} \\ 1 & r \text{ が } p_i \text{ と含意} \end{cases}$$

実験①: 応答に含まれるペルソナは, ユーザー入力に影響されるか?

- 発話と応答からそれぞれペルソナの属性情報を抽出して頻度を比較
- 属性情報にはspaCyの固有表現抽出を利用



- ユーザ発話とシステム応答の属性カテゴリは共起しやすい傾向
- ユーザの発話に関連するペルソナが生成される可能性

実験②: 応答に含まれるペルソナは新しい?

NLI分類器によって応答と既存のペルソナの関係を分類

含意	中立	矛盾	ペルソナ無し
717	4819	489	1479

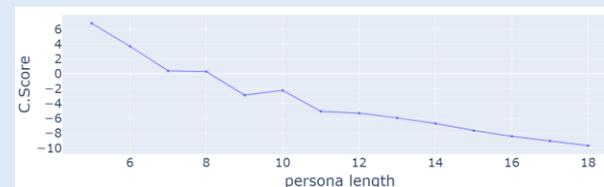
- 半分以上の対話で新しい中立のペルソナが出現
- 応答履歴を用いてペルソナの更新をすることが一貫した対話の実現に必要なといえる

実際の対話例

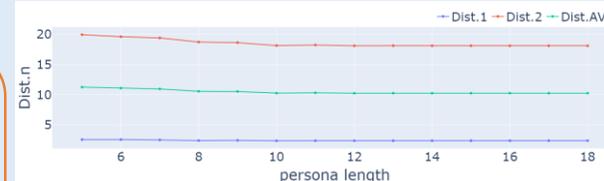
Persona: I am student.
 ユーザー: where do you reside?
 システム: i am in cali, how about you?(caliは地名)
 ユーザー: what is your town famous for?
 提案手法: history and finance. what is your major in college?
 既存手法: a small college in new york

実験③: ペルソナの長さによる応答への影響

矛盾や重複でないペルソナを追加して応答生成を行う



ペルソナの長さによるdistinct-nへの影響



ペルソナの長さによるC.Scoreへの影響

- ペルソナが長くなるほどスコアが下がる傾向
- ペルソナが増えても条件に沿った応答ができていないためスコアが下がっている可能性

実験④: 重複や矛盾が起きた場合にフィルタリングすると結果は良くなるか?

Filter	Dist.AVG(n=1,2)	C.Score	調和平均
BoB	10.691	6.796	4.15486
無し	14.0894	8.1104	5.14737
矛盾	12.2526	22.3411	7.91292
重複	12.4187	13.9311	6.56577
両方	12.2536	16.499	7.03144

- どちらのフィルタもC.Scoreに効果がある
- 両方のフィルタを用いると, 矛盾のみよりC.Scoreが下がってしまう

今後の展望: 他モデルでの実験も予定
 応答履歴によるペルソナ更新の問題設定により適したモデルは何かを考えたい

参考文献

[1] Song et al. BoB: BERT over BERT for training persona-based dialogue models from limited personalized data. In Proc. ACL-IJCNLP, 2021.