

作成: 2022年3月25日

更新: 2022年6月8日

プレゼンテーション資料

北陸先端科学技術大学院大学 先端科学技術研究科
先端科学秘術専攻 博士前期課程 白井研究室 2年

吉田寛太

自由対話システムについて

対話システムとは、人間とコンピューターが自然言語を用いてコミュニケーションするシステムである。

自由対話システムは、コミュニケーションの道具として期待されている。例えば、高齢者の認知症対策としてのコミュニケーションロボットやビデオゲームにおけるNPCとしての役割などが想定される。

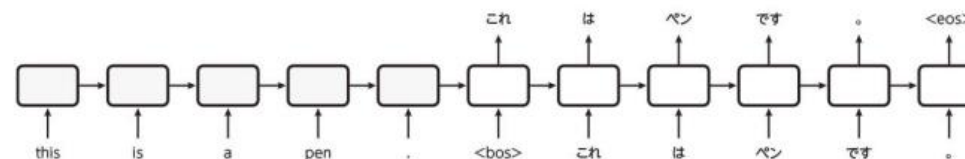


rinna 開発の対話エンジンを搭載した AIキャラクターリアルチャット恋愛ゲーム『プラスリンクス ～キミと繋がる想い～』に登場*1



認知症高齢者に特化した対話ロボ、会話シナリオ開発の工夫 | ニュースイッチ by 日刊工業新聞社*2

近年の自由対話システムに関する研究は、機械翻訳モデルのような深層学習に基づくデータ駆動型のモデルを用い、ソーシャルメディア上の対話を訓練データとして応答を自動生成する手法が主流になっている。



一般的なSeq2Seqモデルによる翻訳タスクの例(英日翻訳)
対話システムでは、発話文を受け取り、応答文を生成する。

*1・・・rinna社「プラスリンクス ～キミと繋がる想い～」 <https://tinyurl.com/2s4b93f8>

*2・・・NECプラットフォームズ「PaPeRoi」 https://www.necplatforms.co.jp/solution/papero_i/index.html

背景

自由対話システムでは、SEQ2SEQのような機械学習モデルを用い、発話から応答を予測するモデルが多く提案されている。

自由対話システムの問題点

Twitter等の様々な話者の発話データを混在した対話データを用いて学習するため
生成された応答に一貫性・個性が欠如しており、ユーザーの対話体験を低下させる事が指摘されている[1]

問題に対する提案されているアプローチ

仮想の応答話者を設定し、応答生成・応答候補選択時にその話者の人格情報(「ペルソナ」と呼ばれる)を参照することで、応答に一貫性・個性を反映する[1, 2, 3].
一方、ペルソナに対応した対話システムが用いるデータセットの作成は多くのクライドワーカーを必要とし、コストがかかる。

| | |
|-----------------|-----------------------------------|
| <i>message</i> | Where do you live now? |
| <i>response</i> | I live in Los Angeles. |
| <i>message</i> | In which city do you live now? |
| <i>response</i> | I live in Madrid. |
| <i>message</i> | In which country do you live now? |
| <i>response</i> | England, you? |
| <i>message</i> | Where were you born? |
| <i>response</i> | I was born in Canada. |
| <i>message</i> | Where are you from? |
| <i>response</i> | England, you? |
| <i>message</i> | Where did you grow up? |
| <i>response</i> | I grew up in Texas. |
| <i>message</i> | How old are you? |
| <i>response</i> | 16 and you? |
| <i>message</i> | What's your age? |
| <i>response</i> | 18. |
| <i>message</i> | What is your major? |
| <i>response</i> | I'm majoring in psychology |
| <i>message</i> | What did you study in college? |
| <i>response</i> | English lit. |

図1: 2500万のTwitter会話データを用いて
4層-SEQ2SEQによる応答予測モデルを学習させた例.
矛盾した応答が目立つ[1].

目的

パーソナリティがタグ付けされた対話データは、その作成コストが大きい。
そのため、Twitterのユーザー情報とツイートのみで構築された対話データだけで性格・一貫性を考慮した自由対話システムを提案する。

応答に性格情報を考慮する

Twitterコーパス(発話Mと応答Rのペアから構成される対話コーパス)から、Rの極性を判定し、
a) 積極的なRから構築される対話データ、b) または消極的なRから構成される対話データ
というa), b)2つの対話データに分割する。このa),b)対話データのみを用いて応答予測モデルを構築することで、応答に個性を反映させる。

ペルソナを自動構築する

文献[2, 3]では、仮想の応答話者を設定し、応答生成にペルソナを参照することで、応答に一貫性・個性が確認された。しかし、ペルソナ情報を記述したデータセットには、どの研究でもクラウドワーカーらを雇用し人手でペルソナ記述している。

一方、TwitterのTWEET, BIOGRAPHY等の情報を特徴量として、ユーザーのプロファイル(性別, 年齢等)を予測するモデルが提案されている[4, 5]。提案手法では、プロフィール予測モデルを用いて、Twitterコーパスに含まれるRの話者のペルソナを自動記述する。

関連研究①

非明示的ペルソナモデル

明示的ペルソナモデル : ペルソナを話者のプロフィール文等のテキストで記述したデータセットを用いるモデル。
非明示的ペルソナモデル: ペルソナを話者IDの埋め込みベクトルやクエリー等のベクトルで表現したモデル。

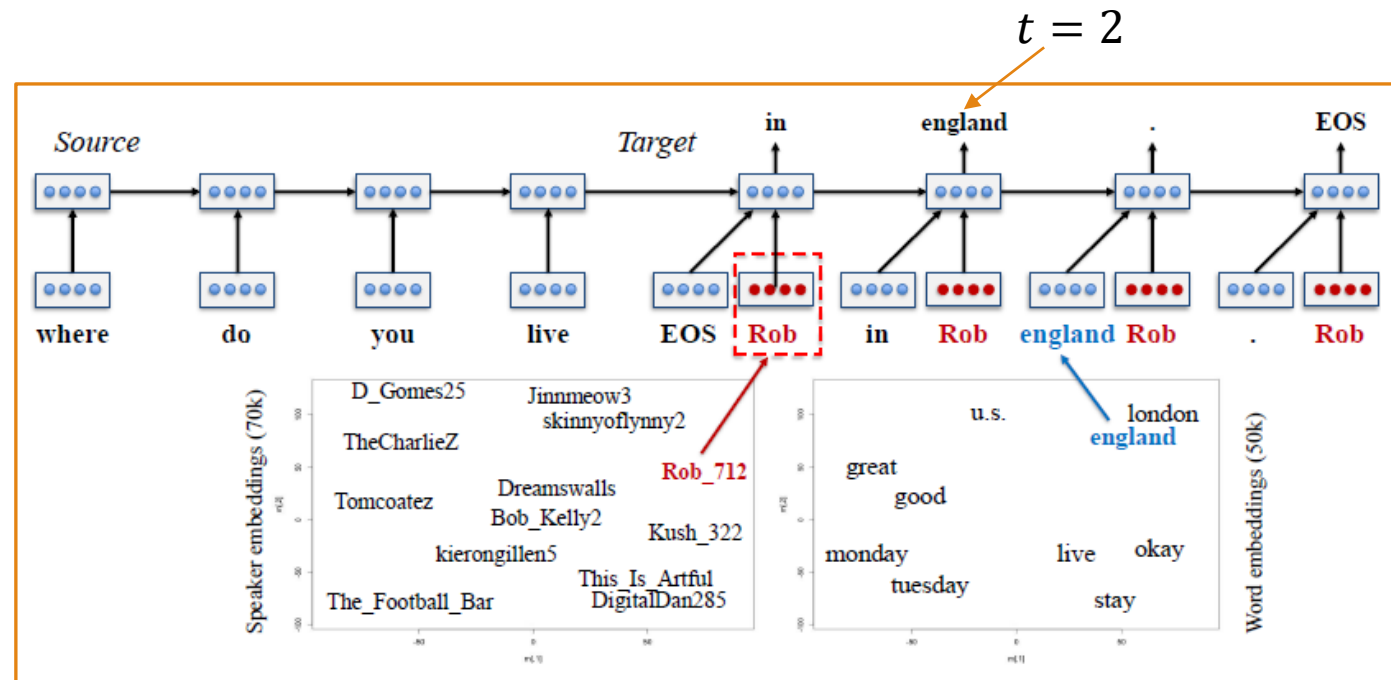
概要

ペルソナを話者IDの埋め込みベクトルとして表現[1]。
応答生成時にターゲット*3とペルソナベクトルを結合することで、生成された応答に一貫性が確認された。

左図の例では話者Robのベクトル Rob を用いて、一貫性のあるin england.を生成している。

Robが教師データ中でEnglandについてよく言及する話者であれば、応答学習時に逆伝播された Rob は、Englandの単語埋め込みベクトルに近くなる。

応答推論時に予測された各ターゲットに対して Rob を結合することで、 $t = 2$ におけるenglandの生成確率は、u.s.やchinaよりも高くなる。



*3...ソースは発話系列, ターゲットは応答系列を指す。

関連研究②

概要

明示的ペルソナモデル. 話者のプロフィール文と対話履歴で構築されたデータセット (Persona-chat Dataset) を用いることで, 応答に一貫性が向上した[2].

以下の2つのモデルを提案.

Ranking Model:

発話ベクトル q からPersona文 p に注意した a を

発話ベクトルに加算 $q^+ = q + a$

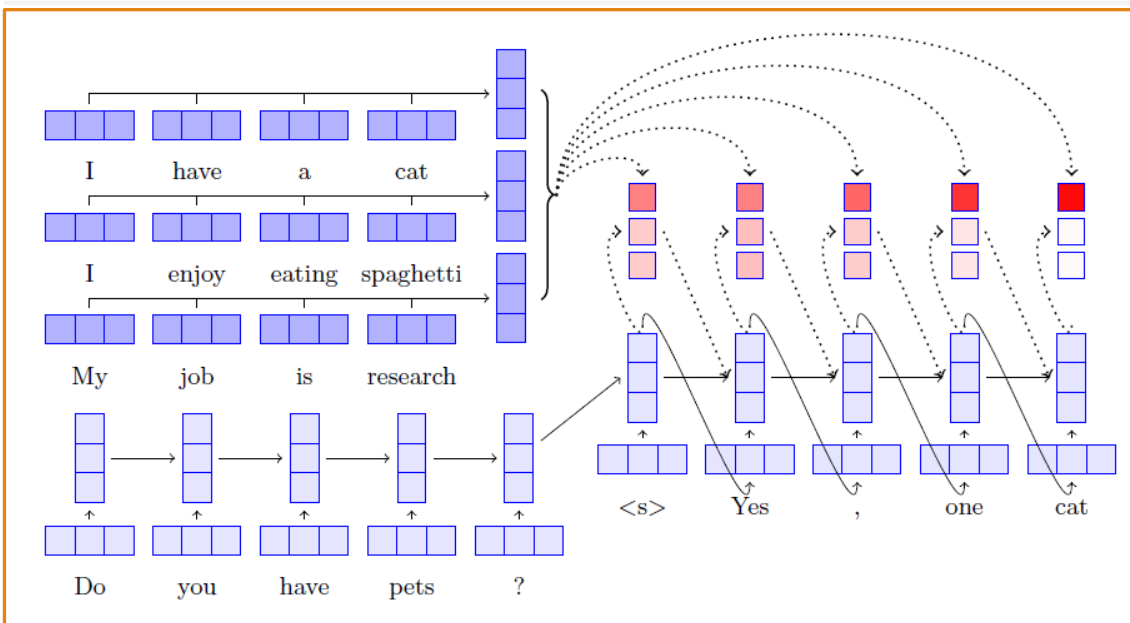
q^+ を対話履歴の応答ベクトルとの類似度を計算し, コサイン類似度が最も大きいものを応答として選択.

Generative Model:

ソースとペルソナを先頭に結合し, LSTMでエンコードした値 h_s を用いてSeq2Seqで応答文を生成. この際, h_s と p を注意する.

明示的ペルソナモデル

| Method | | Persona | | | |
|--------------------------------------|---------|------------|--------------|-------------|------------|
| Model | Profile | Fluency | Engagingness | Consistency | Detection |
| Human | Self | 4.31(1.07) | 4.25(1.06) | 4.36(0.92) | 0.95(0.22) |
| <i>Generative PersonaChat Models</i> | | | | | |
| Seq2Seq | None | 3.17(1.10) | 3.18(1.41) | 2.98(1.45) | 0.51(0.50) |
| Profile Memory | Self | 3.08(1.40) | 3.13(1.39) | 3.14(1.26) | 0.72(0.45) |
| <i>Ranking PersonaChat Models</i> | | | | | |
| KV Memory | None | 3.81(1.14) | 3.88(0.98) | 3.36(1.37) | 0.59(0.49) |
| KV Profile Memory | Self | 3.97(0.94) | 3.50(1.17) | 3.44(1.30) | 0.81(0.39) |



提案手法①

課題

既存手法の多くは、パーソナリティがタグ付けされた対話データを必要とする。その作成コストは高い。

新規性

- Twitterから自動収集した対話データを利用し、応答話者のペルソナを用いることで、個性・一貫性のある対話システムを構築する。
- 対話データには複数の応答話者のユーザーIDとその対話履歴のみが記述されており、人手による対話コーパス構築を必要としない。

データセット

訓練コーパス : 前処理されたTwitter上の対話データから、ツイート(発話)とリプライ(応答)の組を抽出したもの。
ペルソナコーパス : 訓練コーパスからリプライの極性を判定し、以下の操作を行う。

1. 訓練コーパスからリプライの極性を判定し、否定的なツイートには「内向的」、肯定的なツイートには「外向的」のラベルを付与し、内向的もしくは外向的のラベルの付与したツイートの組のみを抽出。
2. プロファイル予測モデルを用いて応答話者のペルソナを構築する。この際、プロファイル予測モデルの信頼性がしきい値より低い話者の対話データに対してはペルソナコーパスから削除する。

提案手法②

方法論

事前学習として、訓練コーパス全体を用いて、ツイート(発話)を受け取るとリプライ(応答)を生成するようなSeq2Seqモデルを学習する。

次に、事前学習したSeq2Seqモデルを初期パラメータとして、ペルソナコーパスを用いてパラメータを再学習することにより、個性・一貫性のある対話システムを構築する。注意機構を用いたSeq2Seqモデルを用い、ペルソナを注意する。

評価

被験者に内向的・外交的な個性を持つ対話システムと対話をしてもらい、「文法」「対話の内容」「性格」についてアンケート調査を行う。

「文法」は対話システムが文法的に正しい発話をしているか、「対話の内容」はシステムとの対話が自然であったか、「性格」は対話システムの個性が内向的・外交的であると感じたかについて、5段階で評価してもらう。

課題

極性判定法の工夫

極性辞書による極性分析は、極性に影響する単語の出現頻度によって計算される。
しかし、文脈によってはネガポジを反転する必要がある。係り受け解析を行うことで反転を判断する。
EX)「怒ってないよ」 $\text{score}(\text{怒っ}) = -0.904882$

個人性の強いペルソナ文を構築できるか？

- Twitterプロフィール予測モデルの多くは、年齢と性別、政治的思考を予測する[4, 5].
- こうした予測によるペルソナの構築は一貫性の向上に寄与する一方、より個人性の強い応答文を生成するには、趣味趣向が記述された個人性の強いペルソナを構築する必要がある。

趣味趣向を予測する範囲

- 趣味趣向を予測する場合はその範囲を限定する必要がある。一方、範囲が狭いほうがより個人性が強くなる。
- 「私はAPEXが好きです」 < 「私はFPSが好きです」 < 「私はゲームが好きです」

REFERENCE

1. Jiwei Li, Michel Galley, Chris Brockett, Georgios Spithourakis, Jianfeng Gao, Bill Dolan. [A Persona-Based Neural Conversation Model](#). Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. pages 994–1003. 2016.
2. Saizheng Zhang, Emily Dinan, Jack Urbanek, Arthur Szlam, Douwe Kiela, Jason Weston. [Personalizing Dialogue Agents: I have a dog, do you have pets too?](#). Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. pages 2204–2213. 2018.
3. Xinchao Xu, Zhibin Gou, Wenquan Wu, Zheng-Yu Niu, Hua Wu, Haifeng Wang, Shihang Wang. [Long Time No See! Open-Domain Conversation with Long-Term Persona Memory](#). Findings of the Association for Computational Linguistics: ACL 2022. pages 2639–2650. 2022.
4. Liu, Yaguang, Lisa Singh and Zeina Mneimneh. [A Comparative Analysis of Classic and Deep Learning Models for Inferring Gender and Age of Twitter Users](#). DeLTA. pages 48–58. 2021
5. Prashanth Vijayaraghavan, Soroush Vosoughi, Deb Roy. [Twitter Demographic Classification Using Deep Multi-modal Multi-task Learning](#). Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. pages 478–483. 2017