

積極的・消極的な性格を持つ 自由対話システムの構築

白井研究室

吉田 寛太

背景(1/3)

- 近年, ユーザーとコンピューターが自然言語で雑談する自由対話システムに関する研究が盛んに行われている
- 人間とコンピュータが自然言語で雑談す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



認知症予防向け対話ロボットの商品^{*1}

^{*1} https://www.necplatforms.co.jp/solution/papero_i/index.html

背景(2/3)

message	Where do you live now?
response	I live in <u>Los Angeles</u> .
message	In which city do you live now?
response	I live in <u>Madrid</u> .
message	In which country do you live now?
response	England, you?
message	Where were you born?
response	I was born in <u>Canada</u> .
message	Where are you from?
response	<u>England</u> , you?
message	Where did you grow up?
response	I grew up in Texas.

Twitter会話データをSEQ2SEQモデルで学習した応答生成モデルによる対話例[Li et al. 2016].

既存研究の問題点

- 深層学習モデルベースの自由対話システム研究では、ソーシャルメディア上の対話から巨大な対話データセットを構築・利用することで、自然な対話を実現する手法が提案されている
- 一方、こうしたデータセットを学習した応答生成モデルでは、生成される応答に個性や一貫性が欠如する問題が指摘されている(左図)
- 訓練データには、ソーシャルメディア上の様々なユーザーの発話が混在しているため、発話に個性や一貫性がないことが原因と考えられる

背景 (3/3)

- 自由対話システム研究では、話者の個人性や背景に関する情報をPersona(ペルソナ)と呼ぶ
- データセットに話者のペルソナ情報をアノテーションすることにより、生成される応答に個性や一貫性を表現する研究がある
[Zhang et al. 2018, Xu et al. 2022]
- これらの手法では、Twitterなどのソーシャルメディア上の大規模な対話データに加え、パーソナリティがタグ付けされた大規模対話データを必要とする
→ クラウドワーカーらを雇用し、対話コーパス中に出現する話者のペルソナ(主にプロフィール文)を手によりアノテーションする
- アノテーション不要なペルソナ対話モデルも存在するが[Li et al. 2016], 応答の一貫性のみに焦点が当てられており、個性を持たせる研究は行われていない

目的

- 本研究では、自由対話システムにおける応答の個性ないしは性格付けの一貫性を保ち、より自然で個性的な対話を実現することを目的とする
- 対話システムの個性として「積極的な性格」と「消極的な性格」の2つを想定する
- Twitterから収集した対話データから積極的(消極的)な発言を選別し、これを訓練データとすることで、積極的(消極的)な性格を持つ対話システムを実現する

関連研究

□ Li et al. 2016

Encoder-Decoderモデルで応答を生成する際, Decoderの各時刻の入力に話者ベクトルを加算することで, 話者の特性を応答に反映させた

□ Zhang et al. 2018

クラウドワーカーらにプロフィール文によって設定された人物を演じながらテキストチャットを行わせ, その対話履歴から応答生成モデルを学習することで対話システムの応答に一貫性を与えた

□ Xu et al. 2022

ペルソナと対話履歴から構成されたデータセットに対し, 発話が既知のペルソナを参照しているか, 新たに記録すべき未知のペルソナを含むかをラベル付けした

本研究の特色

- 既存手法の多くは、ペルソナがタグ付けされた対話データセットを必要とする.
- 本研究では、人手によって付与されたペルソナを必要とせずに積極的・消極的といった個性を持つ自由対話システムを構築する

提案手法(1/2)

対話モデル

- 自由対話システムは系列変換モデル(Seq2Seqモデル)で実現できる
- Seq2SeqモデルとしてTransformer[Polosukhin et al. 2017]を用いる
- Twitter上から収集した発話と応答の組を訓練データとし, Transformerを学習する

提案手法(2/2)

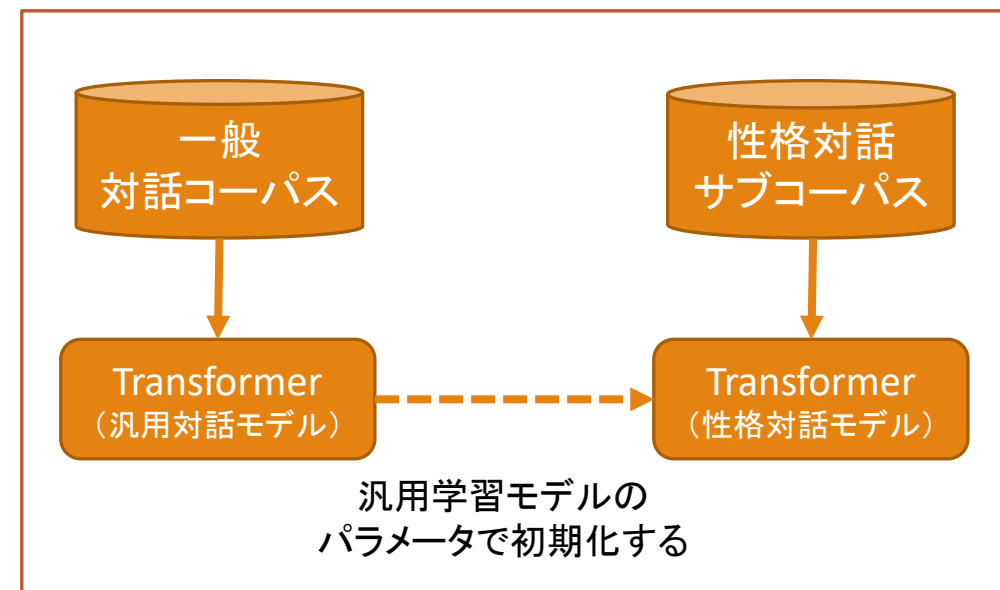
二段階学習

□ 汎用対話モデルの学習

- 比較的大規模な対話データから汎用的な自由対話システムを学習する

□ 性格対話モデル(積極的対話モデル, 消極的対話モデル)の学習

- 消極的な応答, 積極的な応答を抜粋した対話コーパスを作成する
- これを性格対話サブコーパス(積極的対話サブコーパス, 消極的対話サブコーパス)と呼ぶ
- 汎用対話モデルを初期値とし, 性格対話サブコーパスを用いてTransformerを再学習する



訓練データの構築(1/2)

Twitter対話コーパス

- Twitterにおけるツイートとリプライの連鎖を擬似対話として収集したコーパス[関田 2020]を使用する
- ユーザーID, ハッシュタグの除去といった前処理を行う
- 対話(ツイートの連鎖)を2つの発話の組に分解し, 発話と応答の組から構成されるデータセットを構築する

訓練データの構築(2/2)

□ 性格対話サブコーパス

- Twitter対話コーパスの応答の極性を判定し, 肯定的な応答は積極的, 否定的な応答は消極的とみなす
- 応答が積極的または消極的な発話と応答の組を抜粋し, 積極的対話サブコーパス, 消極的対話サブコーパスを作成する

□ 一般対話サブコーパス

- 極性判定で中立と判定された発話・応答ペア
- 汎用対話モデルの学習に用いる

ツイートの極性判定

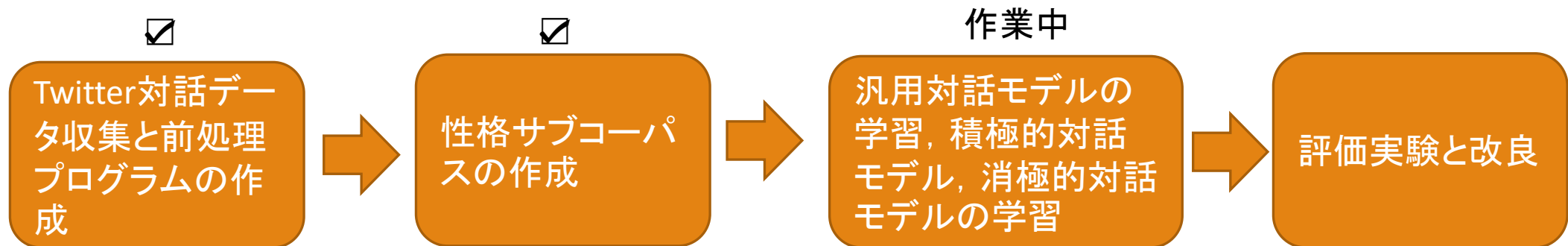
以下の手続きでツイートの極性を判定する.

- ツイートを形態素解析する
- 日本語評価極性辞書[小林ら 2005]を用いて, ツイート中の単語の極性スコアを求める
 - 単語の極性スコアは-1または1の値
- ツイート中の単語の極性スコアを平均を計算し, それが0以上のときは肯定的, 0未満のとき否定的と判定する
- 否定表現の処理
 - 「このゲームは楽しくなかった」という文では, 「楽しい」の極性スコアは1だが、否定表現により文全体の極性は否定的である
 - 文節の係り受け解析を行い、否定表現に係る単語の極性スコアを反転する

評価

- 被験者に対話システムとチャット形式で対話してもらう
- 以下の3つの項目について5段階評価のアンケート調査を行う
 - 文法評価：システム応答は日本語として文法に破綻がないか
 - 対話内容：システム応答の内容は適切であったか
 - 個性評価：システムの性格を積極的または消極的と感じたか
- 2つの対話システムを比較する
 - 汎用対話モデル(ベースライン)
 - 積極的対話モデルまたは消極的対話モデル(提案手法)

現在までの進捗



作成したデータセット

- 右表にデータセットの詳細を示す
 - ここでターンとは、発話-応答のペアを指す
- 積極的対話サブコーパス、消極的対話サブコーパスは一般対話サブコーパス(中立の応答)に比べて小規模である
- Twitterにより収集したコーパスを用いた類似研究 [Li et al. 2016] に比べ、本研究のデータセットのサイズは小規模である
- 各サブコーパスの14%を開発データ, 6%をテストデータとして分割した

極性限定対話コーパス	ターン数
積極的対話サブコーパス	18,135
消極的対話サブコーパス	36,425
一般対話サブコーパス	588,162
Twitter Persona Dataset [Li et al. 2016]	24,725,711

対話例

一般対話モデルによる対話例

Source : 了解ですなーなるみんとか呼ばれるので好きに呼んでください! タメにしょー

Target : やったー!

Response: ぜひ仲良くしてください

Source : みこちゃん呼びしますね! はるって言うのでお好きなように呼んでください!

Target : よかったら敬語ポイして話ませんか?

Response: タメで話しましょう

Source : 寿司屋巡りでもあるよねwww

Target : もう、元々の趣旨が残ってないwww

Response: それはそれで、.....
.....
...

Source : 万事屋も同じじゃねえのか?

Target : 祝い酒とかはしないけど騒いだりするわ

Response: それはない

対話例の作成には
一般対話サブコーパスの開発データを利用した.

Source ... データセット中の発話文

Target ... データセット中の発話に対する応答

Response ... 対話モデルが生成した応答文.

対話例

積極的対話モデルによる対話例

Source : はじめまして!ありがとうございますぜひ繋がりたいですフォロー失礼します!

Target : 早速ですが、何とお呼びしたら良いですか??

Response: 呼びタメ大歓迎なので仲良くしてください

Source : わいはラジオも無駄使いだけ聞いてました笑
あ!ガンダムみてますよ!

Target : ダブルオーは、絵綺麗だからいいよ!

Response: みんなで楽しみましょう.....

Source : おはようございますベランダ菜園良いですねー

Target : ただ虫がくる・・・

Response: やりましょう

Source : すみません意味のない絵文字をよく使うので
クセで使ってしまいました..

Target : 良いアクセントでしたエビw

Response: 次は、、、、、、、、、、、、、、、、、、、、、、

消極的対話モデルによる対話例

Source : わあ、美味しそうなトンカツ.....エビフライだと!?

Target : お仕事でないとちょっと厳しいです

Response: 一口で食べたら太るぞ

Source : 愛知だから今回の富士急は苦勞するんだよw

Target : 富士急アクセス悪いからねw

Response: じゃ、俺は多分無理

Source : あの旅は素晴らしかったわ、にったんwww

Target : いやー呆れたけどよく頑張ったw

Response: その調子で死にたいけど怖くてない

Source : 今のところそんなに暑くなかったけど、これからだねー

Target : そっちはそんなに暑くないんや、こっちはもう地獄の蒸し暑さやでー。

Response: そうなんだね、私も大変だったよ

考察

- 汎用対話モデルではある程度自然な応答が生成可能
- 性格対話モデルでは一部発話に対する応答内容が不自然である
 - Twitter会話コーパスを用いた先行研究に比べデータ量が少ないことが原因？
- 汎用対話モデル，性格対話モデルの両方のモデルで不自然な語の連続が確認される
 - (例) こんにちは。。。。。。。。。。。。。。。
- ハイパーパラメータ探索をする必要があるか
 - TransformerのEncoder/Decoderのレイヤー数，Dropout値，学習率

今後の予定

□ 性格サブコーパスの評価

- 少量のテストデータを人手作成し、極性判定手法の正解率を測り、それがある程度高いこと(性格サブコーパスの品質が十分に良いこと)を確認する

□ データセットの拡張

- Twitterからの新たな発話対の収集や言い換えによるデータ拡張などの手法を検討する

□ 事前学習モデルの利用

- 下記のような事前学習モデルを性格サブコーパスを用いてファインチューニングする
 - ・ Transformers事前学習対話モデル
 - ・ BART日本語Pretrainedモデル[田中ら 2021]

□ 他の性格や個性を持つ対話システムの実現

- ユーザのプロフィール(職業, 出身地)を自動推定することで, ペルソナを自動的に付与する方法を検討する

参考文献

- 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. pp 2204–2213. 2018.
- 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. pp 994–1003. 2016.
- 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. pp 2639–2650. 2022.
- 小林のぞみ, 乾健太郎, 松本裕治, 立石健二, 福島俊一. 意見抽出のための評価表現の収集. 自然言語処理, Vol.12, No.3, pp.203–222, 2005.

参考文献

- 関田 崇宏. マイクロブログからの対話コーパスの自動構築. 北陸先端科学技術大学院大学 修士論文, 2020.
- 田中佑, 村脇有吾, 河原大輔, 黒橋禎夫: 日本語Wikipediaの編集履歴に基づく入力誤りデータセットと訂正システムの改良, 言語処理学会第27回年次大会, 2021.

対話例

積極的対話モデルによる対話例

Source : そうだね頑張ろうね華、気をつけて行ってらっしゃい

Target : 華休憩はいったよ

Response: **ありがとうございます**

Source : ニートになった元占い師です。こちらこそありがとうございます!

Target : ありがたくぼっぼちゃんはいいただきますね

Response: **ありがとうございます**

Source : 12秒くらい考えた

Target : かわいい

Response: **それは草**

Source : おはようございます!こちらは霧が濃く、寒い朝です良い木曜日をお過ごし下さいませ

Target : 楽しい金曜日をお過ごしください

Response: **楽しい木曜日を!**

Source : 多いですね私の住まいは田舎なので夜バイクで走っているとよく虫にぶつかります

Target : 帰ったら即風呂入りたいですね

Response: **私もです**

Source : cd買ったよ!!

Target : 自引き出来た?

Response: **よーしよしよしよしよしよし**

Source : 年末なごやのなにかの現場に合わせてやりたいね!!!たのしみ!!!

Target : 手羽センカウントダウンあればいいのに

Response: **一緒に、いこう**

Source : すごく昔からの知り合い感やばかったよな。まじで楽しかった.....

Target : 会ったことあって2回目のような居心地の良さよ。

Response: **いれば良かった**

消極的対話モデルによる対話例

Source : 過去問やればいけるっしょ!!最悪追試あるし
ね笑笑

Target : 最悪追試あるしね

Response: 今はも
う、、、、、、、、、、、、、、、、、、
、、、、、、、、、、、、、、、、、、

Source : 走ってるうちは大丈夫ですへへへ
Target : 問題はスタートか笑
Response: 私は今からは以上に生きています