B-02

特定人物を模倣したチャットボット作成システムの開発

Development of the Chatbot Creating System to Imitate a Specific a Person

三木　康太[1]　宮部　真衣[2]　吉野　孝[3]
Kota Miki　Mai Miyabe　Takashi Yoshino

1. はじめに

近年, 質問応答や案内等を行うタスク指向の対話システム以外に, 雑談を主目的とした対話システムが増加している. そのような対話システムでは会話相手を楽しませるような発話が重要であり，発話の面白さや自然さに着目した研究が盛んに行われている.

雑談目的の対話システムで必要な要素の一つに，個性が存在する．個性ある応答を行うことで，親近感や人間らしさを感じやすくなり，より楽しい会話を行えると考えられる．しかし，個性ある応答を人手で作成することはコストが高く，近年では個性ある応答の自動生成に関する研究が存在する [1][2]．だが，応答を自動生成した場合，発話内容を制御することが困難である．2016 年には Microsoft のチャットボット「Tay」が不適切な発言をしたことで問題になり，公開が停止された事例が存在する[*1]．また，自動生成した応答は対応できるトピックが広いものの，人手で作成した応答に比べ，応答の品質は不十分である．

そこで本研究では，既存のチャットボットに個性の追加反映機構を付加することで，新たなチャットボットを作成するシステムを開発した．本システムでは，既存のチャットボットの応答に対し，個性の追加反映を行い，個性ある応答を行えるチャットボットを作成する．既存の高品質なチャットボットと連携することで，新たな個性を備えたチャットボットを容易に作成することが可能となる．

本稿では Twitter のリプライに対し，個性を付加した応答を作成し，個性の付加が可能であることを示した.

2. 関連研究

個性を備えた応答生成に関する研究の一つに，赤間らの研究 [1] がある．Twitter の大規模な対話データを使い，seq2seq による応答生成モデルを作成し，少数のスタイル付き対話データを転移学習を利用することで応答生成モデルを作成している．本研究と使用データや手法の面で類似しているものの，本研究は応答生成を伴わない個性反映である.

また，Li らの研究 [2] では，人物の情報をベクトル空間に埋め込むことにより，特定の人物らしい応答生成をしている．これは赤間らの研究と同様に応答生成を行う研究であるため，本研究とは目的の面で異なる.

濱田らの研究 [3] では，模倣対象の人物ごとにベクトルを作成し，発話時に任意の人物のベクトルに寄せた応答の生成と文章の書き換えによる応答の生成手法について述べている．文章の書き換えは本研究と類似しているものの，既存のチャットボットと連携することは難しく，専用のモデルが必要とされる.

表 1: 個性化した発言の生成例

入力	出力
そうです	そうですね
改めて、宜しく	改めて、宜しくね
どこか行きたいとかあるか？	どこか行きたいとかあるか??

3. 提案システム

本システムでは個性の追加反映を行うために Sutskever らの seq2seq[4] を使用する．模倣対象の発言と無個性化した発言のペアを学習することで，模倣対象の文体を真似た発言を seq2seq により生成する.

3.1 モデルの学習

本システムのモデルは転移学習を利用し作成する．多くの場合，特定人物の発言は利用できるデータ量が限られており，深層学習に必要なデータ量を集めることは困難である．そこで，関連するデータで事前の学習を行った後，本来のデータで学習を行う手法 (転移学習) を使用する．この手法は，赤間らによる研究 [1] において，応答生成に有効であると示されている.

事前の学習として，無個性な発言から個性的な発言を生成するモデル (以下，事前学習モデル) を作成する．学習に用いるデータは無個性な発言と個性的な発言のペアである．個性的な発言とは，語尾や単語に人物特有の情報が含まれているテキストのことであり，無個性な発言は個性的な発言を 3.2 節の処理を行うことで作成する．ただし，作成したペアが全く同一であると適切な学習が行えないため，対になるテキストで一単語以上の相違がある場合のみ学習データとして使用する．また，学習データは形態素解器 Mecab[*2] による分かち書きを行う．上記の手順で作成したペアを seq2seq による学習に使用することで，無個性な文から個性的な文を生成するモデルを作成する．表 1 に事前学習モデルによる文章生成例を示す．例では Twitter のデータを使用したモデルによる文章生成の結果が示されており，感嘆符や終助詞が付加されていることが分かる.

次に，無個性な発言から特定の人物らしい発言を生成するモデル (以下，特定人物モデル) を作成する．事前学習モデルの学習データと同様に，模倣対象から獲得したテキストを無個性化処理を行うことで無個性発言，個性発言のペアを作成する．このペアを事前学習モデルに転移学習することにより，無個性発言から個性発言を生成するモデルか

[1] 和歌山大学大学院システム工学研究科，Graduate School of Systems Engineering, Wakayama University
[2] 公立諏訪東京理科大学工学部，Faculty of Engineering, Suwa University of Science
[3] 和歌山大学システム工学部，Faculty of Systems Engineering, Wakayama University

[*1] マイクロソフトの AI「Tay」が一時復活、また暴言: https://japan.cnet.com/article/35080422/

[*2] http://taku910.github.io/mecab/

表 2: 削除対象の形態素

・	…	...	ー	ーー	～	！	♪
w	ｗｗ	ｗｗｗ	ｗｗｗｗｗｗｗｗｗ				

表 3: 非削除対象の形態素

いる	う	おく	か	がる	させる	せる
た	たい	だ	っけ	てる	です	とる
な	ない	へん	べき	ほしい	まい	みる
や	らしい	られる	れる	ん	欲しい	

ら，無個性発言から特定の人物らしい発言を生成するモデルに変更する．本システムでは転移学習の手法として事前学習モデルを初期値とした転移学習を利用した．

3.2 無個性化処理

モデルの学習に使用するために，個性的なテキストを無個性なテキストに変換する処理が必要である．モデルの学習には大量のデータが要求されるが，ある個性が反映されたテキストと別の個性が反映されたテキストのペアを大量に収集することは難しい．だが，個性的なテキストを無個性にすることは容易であるため，何らかの個性が反映されたテキストを無個性にすることで学習用のデータとする．

無個性化処理は形態素をルールベースで削除することで行う．文章中における個性を表す要素は先行研究 [5] でいくつか述べられているが，本研究では単語と語尾が特に個性を表すと考えた．そこで，個性を表すと考えられる単語と語尾を削除することで，無個性な文章に変換する．個性があると考えられるテキストを Mecab により形態素解析し，絵文字，フィラー，未知語および表 2 に該当する形態素を削除する．

さらに，述語機能部 (語尾) に含まれる単語は膨大な量が考えられるため，表 3 の除去すると意味が変わると考えられる形態素以外を削除する．

4. 実験

本システムが個性反映を適切に行えるかを検証するために，ランダムな発話に個性を付加する実験を行った．

4.1 学習データ

事前学習モデルの学習データは Twitter から取得したツイートを用いた．2017 年 4 月から 2018 年 7 月までのランダムに取得したツイート 217 万件に対し，URL や日本語以外の言語を除去する処理を行い，ツイートをランダムに 20 万件抽出した．また，長い文章は対話システムの応答に適さないと考えられるため，ツイートは 64 単語以下に限定した．

特定人物モデルを学習するデータは非実在のキャラクタ 3 名のツイートを用いた．模倣対象は，文体に特徴がある人物から学習した方が精度評価が容易であると考えため，以下の条件で決定した．

(1) ツイッターアカウントを所有

(2) ツイートを見ただけで人物特定が可能

(3) 非実在のキャラクタ

表 4: 実験に使用するツイートの例

アカウント名	ツイート内容
funassyi	(。゜▽゜) 梨ドック！ (。゜▽゜) コッチミンナ かわゆすなっしー♪ヾ (。゜▽゜) ノ
55_kumamon	今日はアップで、おはくま☆ おやくま～ … ☆ おはくま～！今日も 1 日よろしくま☆
kirimi_sanrio	おはよう…まだねむい… ぺこり…KIRIMI ちゃん. です！ よいしょ…っと…！ふ～

上記の条件を満たし，フォロワー数が多い「funassyi」「55_kumamon」「kirimi_sanrio」の 3 アカウントが発信したツイートを利用する．上記の 3 アカウントから取得したツイートから，リツイートやリプライを除外した結果，取得できたツイート数はそれぞれ 1685，793，1842 である．表 4 にそれぞれのツイート例を示す．

4.2 モデル設定

モデルの学習では 2 層の GRU を使用し，単語ベクトルが 128 次元，隠れ層が 256 次元，バッチサイズを 16 としてミニバッチ学習を行った．語彙は事前学習データに存在する単語の出現頻度上位 50,000 単語を使用した．エポック数は事前学習モデルで 50,000 回，特定人物モデルでは 1,000 回の固定値である．また，エンコーダ，デコーダ共にアテンションモデルを用いた．

4.3 テストデータ

実験に使用するデータはランダムに抽出したリプライツイートを利用する．本システムはチャットボットの応答に個性の追加を行うシステムであるため，しばしばチャットボットの応答として利用されているツイートを実験に使用する．4.1 節で使用した条件に加え，別のツイートに対するリプライをしているツイート (リプライツイート) に限定し，ランダムに 30 件抽出した．抽出したツイートに対し無個性化処理を行い，特定人物モデルへの入力とした．

4.4 結果

生成結果の一例を表 5 に示す．表 4 と比較すると，funassyi 特有の顔文字や，kirimi_sanrio コーパスで頻出の三点リーダが反映されており，適切に個性が反映されていることが分かる．一方で 55_kumamon コーパスでは,「くま～」という語尾が頻出しているものの，本システムが生成した応答にこの語尾が反映されることは少数であった．このことから，funassyi および kirimi_sanrio コーパスを使うと適切な個性反映がなされるが，55_kumamon コーパスでは不十分な個性反映となる傾向が分かる．

適切な個性反映が行われたかを特有の単語や語尾が含まれているかという基準で人手で分類した結果，適切な個性反映は 78 件 (86.6%)，不適切な個性反映は 12 件 (13.4%) であった．一方で，個性は反映されたものの，生成された応答の意味が元の応答から変化した場合も存在し，そのような不適切な応答の生成例を表 6 に示す．

表 5: システムによる応答生成の例

キャラクタ	入力	出力
funassyi	ほい．凄い凄い 改めて、宜しく やだあ	ほいー (。゜▽゜) 凄いー凄いー凄いなっしー (。゜▽゜) そして宜しくー宜しくー や だ あ (。゜▽゜) あ だ
55_kumamon	ほい．凄い凄い 改めて、宜しく やだあ	… 凄い凄い！ 凄い！ 改めて、ハッピー … や だ ～ ・ ・ ・
krimi_sanrio	ほい．凄い凄い 改めて、宜しく やだあ	ほ～い…！凄い…！ 改めて 、 宜しく ね … やだ……きり…

表 6: 不適切な応答生成の例

キャラクタ	入力	出力
funassyi	ワイにはアドバンテージがない 私が化けたらお兄さんのせいだから 面倒くさいから抜糸なし	(。゜▽゜) 冗談がにがない (。゜▽゜) 私の夜だからなんか同じのだから (。゜▽゜) ごろ
55_kumamon	権力に潰される 面倒くさいから抜糸なし お煎餅あるんだ	数字に潰されるよ～！ … 復興！ お … ひ～だ～☆
kirimi_sanrio	いや誘った LINE 返せ 面倒くさいから抜糸なし お煎餅あるんだ	いや…誘った…？今… から なし…！ お…あるんだよ…

5. 考察

不適切な応答が生成された原因は，個性の反映が不十分である場合，および文法が不適切である場合の 2 種類が存在する．個性の反映が不十分である原因は，学習データの内容にあると考えられる．例えば，55_kumamon のツイートでは「おはくま～！」「いってらっしゃいだモン！」のように「くま」「モン」という語尾が使用される傾向がある．しかし，そのようなテキストは正しい形態素解析が行えないため，無個性化処理で語尾を削除することが不可能である．つまり，無個性文章と個性文章のペアが作成できないため，理想的な学習データが作成できない．この問題を解決するためには，形態素解析を利用しない無個性化処理が必要であると考えられる．

文法が不適切であるという問題には複数の原因が考えられる．一つは学習語彙の問題である．4.2 節に示したとおり，今回の学習では出現頻度上位 50,000 単語を利用したが，固有名詞のような低頻度で出現する単語は含められない．この結果，出現頻度が低い単語は出力ができないため文法の誤りが発生する．他の文法誤りの原因として，事前学習に利用したデータの質が問題であると考えられる．例えば，「おはくま～！」という文章は無個性化処理で「おはくま」と変換され，個性が残されたままとなる．このように無個性化処理が不十分である場合が存在したことが，非文が生成される一因であると考えられる．また，形態素解析に起因する不適切な分かち書きが存在したことも非文生成の原因であると考えられる．

6. おわりに

本研究では，チャットボットの応答に対し，個性の追加反映を行うことで個性あるチャットボットを作成するシステムを開発した．実験では，ツイートを元にした応答に対し，86.6%の割合で任意の個性を反映することが可能であることを示した．一方で，実験では明確な文章上の個性を持つキャラクタを用いたため，模倣対象次第では適切に個性反映が行えない可能性がある．

今後の課題として，対象キャラクタにより模倣精度が変わるため，より適切な応答を生成できる手法の開発が必要であると考えられる．

参考文献

[1] 赤間 怜奈, 稲田 和明, 小林 颯介, 佐藤 祥多, 乾 健太郎：転移学習を用いた対話応答のスタイル制御，言語処理学会 第 23 回年次大会 発表論文集, pp. 338-341 (2017).

[2] Jiwei Li, Michel Galley, Chris Brockett, Georgios Spithourakis, Jianfeng Gao, and Bill Dolan. A persona-based neural conversation model. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, pp.994-1003(2016).

[3] 濱田 晃一, 藤川 和樹, 小林 颯介, 菊池 悠太, 海野 裕也, 土田 正明：対話返答生成における個性の追加反映, 自然言語処理研究会報告, Vol.2017-NL-232, No.12, pp. 1 - 7 (2017).

[4] Ilya Sutskever, Oriol Vinyals, and Quoc V Le. Sequence to sequence learning with neural networks. In Advances in neural information processing systems, pp.3104-3112(2014).

[5] 宮崎 千明，平野 徹，東中 竜一郎，牧野 俊朗，松尾 義博，佐藤 理史：文節機能部の確率的書き換えによる言語表現のキャラクタ性変換，人工知能学会論文誌，Vol.31, No.1(2016).