

対話返答生成における個性の追加反映

濱田 晃一^{1,a)} 藤川 和樹^{1,b)} 小林 颯介^{2,c)} 菊池 悠太^{2,d)} 海野 裕也^{2,e)} 土田 正明^{1,f)}

概要: 個性を反映した返答生成は、様々な個性のキャラクタを用意することで多様な返答を可能にしたり応答の一貫性を確保できる点で近年注目されている。一方で、それぞれの返答個性を学習させるための対話例の準備が高コストであるという問題がある。本研究では、深層学習を用いた返答文生成モデルに対して、特定の返答個性を追加反映する手法の提案及び比較評価を行う。さらに、返答文生成に用いる個性表現ベクトルの演算による生成文の変化について分析を行う。

HAMADA KOICHI^{1,a)} FUJIKAWA KAZUKI^{1,b)} KOBAYASHI SOSUKE^{2,c)} KIKUCHI YUTA^{2,d)}
UNNO YUYA^{2,e)} TSUCHIDA MASA AKI^{1,f)}

1. はじめに

近年、チャットによるコミュニケーションツールが普及し、それに伴って対話エージェントを用いた様々なサービスの開発が行われている。こういったサービスでは、対話エージェントにキャラクタ性を出すための個性を与えることで、親しみやすさやエンターテインメント性を向上することも重要と考えられる。

本研究では、雑談の返答生成において、キャラクタ設定としての個性を追加反映できる雑談対話エージェントを提案し、評価した結果を報告する。我々は、自然な返答生成と個性のモデル化のために、Liらの手法 [6] を用いる。Liらの手法では、クエリ(元の発話)とレスポンス(クエリに対する返答)とレスポンスを返した返答ユーザ(レスポンスの発話者で固有のIDを持つ)の3つ組を訓練データとして、クエリに対する返答ユーザの個性を反映した返答生成のモデルを学習する。ユーザをベクトルとして表現し、返答ユーザに条件づけられた返答生成モデルを学習することによって、そのベクトルにユーザの個性の情報が埋め込まれるとみなせる。そのため、本研究ではこのベクトルを個性表現ベクトルと呼ぶ。しかしながら、Liらの手法で作ら

れる個性表現ベクトルは訓練データに存在するユーザに限られ、このままでは任意の個性により構成された返答生成を行うことができない。

本研究では、この課題の解決方法として、元の個性表現ベクトルによる生成レスポンスの基本的な内容を変えずに、目的とする個性を追加反映するように個性表現ベクトルを変換する手法 (*Transformation, Interpolation*) を検討する。また、別のアプローチとして、生成レスポンスに対して、内容を変えずに個性が追加反映されるように書き換えるモデルを適用する手法 (*Translation*) を検討する。*Translation* は、クエリに対するレスポンス生成ではなくなるため、自然な返答にならない可能性は高まるが、内容を変えずに個性を追加反映する狙いには直接的である。実験では顔文字多用 (*Kaomoji*)、関西弁 (*Kansai*)、熱血漢 (*Aniki*) の追加付与を題材に主観評価を行ない、生成事例を詳細に確認してそれぞれの手法の特性について分析する。

2. 関連研究

2.1 雑談の返答

雑談対話の返答のアプローチには、大きく分けて、用例ベース ([1], [11], [12] など) と生成ベース ([5], [11], [15], [16] など) がある。

用例ベースは、スコアリングモデルにより、クエリに対するレスポンスを候補集合から検索して提示する。この手法はウェブテキスト等を使うことで、様々な話題に合わせてレスポンスを提示できるという利点がある。一方で、クエリに対して自然なレスポンスにならないことがある。例

¹ 株式会社ディー・エヌ・エー
² 株式会社 Preferred Networks
a) koichi.hamada@dena.com
b) kazuki.fujikawa@dena.com
c) sosk@preferred.jp
d) kikuchi@preferred.jp
e) unno@preferred.jp
f) masaaki.tsuchida@dena.com

えば、ある映画のストーリーに関するクエリに対して、その映画の主演の詳細情報を返してしまうなどである。すなわち、様々な話題をカバーできる柔軟性はあるが、対話としての自然さの担保は難しいという問題がある。

生成ベースは、訓練データから文の生成確率を学習し、学習された分布に基づいて文を生成させる手法全般を指す。一般的には単語ごとに生成確率を学習し、単語の列として文を生成する。単語列を自由に生成させる自由度があるため、事例集合から選択する用例ベースの手法と違って、未知の文を生成することがある。その反面、非文などの不適格な文を返答する可能性も高まる。

言語的に自然なレスポンスの生成は、深層学習ベースの言語モデルの近年の進展により徐々に可能性が見えてきている [8], [13]。先駆的な研究は Vinyals ら [15] であり、クエリを入力にレスポンスを出力する Encoder-Decoder を学習することで、返答としても言語としても自然なレスポンスを生成できることが示されている。Li ら [5] は、単純に Encoder-Decoder を学習すると「I don't know.」などクエリに関係なく通じるレスポンスが生成される傾向がある問題に対して、クエリとレスポンスの相互情報量を大きくするように生成モデルを学習する手法を提案している。Zhou ら [17] は、Encoder-Decoder の学習において、訓練データ中で同じクエリに対して異なるレスポンスがある現象を捉えた学習を行うために、クエリ \mathbf{x} からレスポンス \mathbf{y} の生成が、隠れたメカニズム m を経由し $p(\mathbf{y}|\mathbf{x}) = \sum_{m_i} p(m_i|\mathbf{x})p(\mathbf{y}|\mathbf{x}, m_i)$ で生成されていると考え、学習する手法を提案している。

2.2 個性の反映

宮崎ら [20] がまとめている通り、個性の反映に関連する先行研究では、話者の性格 [7], 印象 [22], 人物属性 (性別, 年代, 居住地等) [19], 特定の話者らしさ [10], [21] などが取り込まれている。

Li ら [6] は、対話例からクエリに対するレスポンスの生成とユーザ毎のレスポンスの違いを表現する表現ベクトルを学習することで、ユーザ自身を「個性」として取り込むことができる手法を提案している。様々な個性の違いをその個性の表現ベクトルの違いとして扱い、ベースとなる Encoder-Decoder は共通化することで、対話としての自然さと個性の扱いが可能になると考え、我々のベースモデルとして採用している。

個性の反映とは異なるが、Hu ら [2] は、感情や時制を反映したテキストの生成手法を提案している。感情や時制を属性値として表現し、Encoder の出力と付与したい属性値の表現ベクトルを Decoder への入力に用いている。属性値の表現ベクトルは、Discriminator が、Decoder の生成サンプルからその属性値を予測できるように、Encoder-Decoder, Discriminator と合わせて学習している。ユーザの個性を属

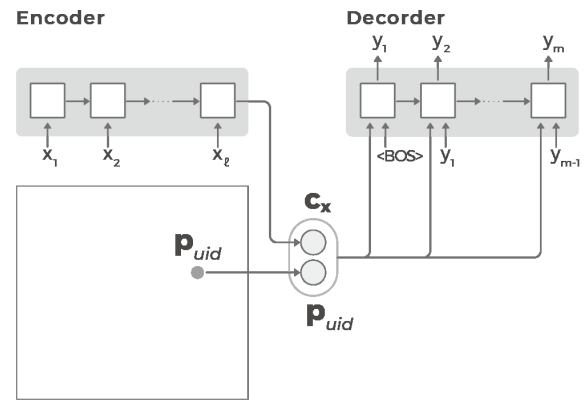


図 1 Speaker Model

性値と考えると個性の反映として利用することができる。

3. 返答個性の追加反映手法

本節では、返答文生成において、特定の個性を追加反映する手法についての解説を行う。最初に、ベースとなる Li ら [6] の Speaker Model について説明する。次に、所望のキャラクター性を得るために、Speaker Model によって得られた個性表現ベクトルを利用して個性追加反映の手法について説明する。最後に、生成されたレスポンスを書き換えることにより個性追加反映を行う手法について説明する。

3.1 Speaker Model [6]

Speaker Model では、返答ユーザ毎の個性表現ベクトルと Encoder の出力を Decoder の入力として、返答生成の Encoder-Decoder の学習を行う (図 1)。具体的には、式 1 の条件付き対数尤度最大化でパラメタを学習する。

$$\max_{\theta, \mathbf{P}} \sum_{(\mathbf{x}, \mathbf{y}, uid) \in D} \log p_{\theta}(\mathbf{y} | \mathbf{x}, \mathbf{p}_{uid}), \quad (1)$$

ここで、 $\mathbf{x} = (x_1, \dots, x_l)$ は可変長のクエリの単語系列、 $\mathbf{y} = (y_1, \dots, y_m)$ は可変長のレスポンスの単語系列、 uid は返答ユーザの ID、 \mathbf{p}_{uid} は返答ユーザの個性を表す個性表現ベクトルである。学習するパラメタは、返答ユーザの個性表現ベクトルの集合 \mathbf{P} と Encoder と Decoder のそれぞれの LSTM のパラメタと Encoder-Decoder で共通の単語の埋め込みベクトルの集合である。LSTM と単語の埋め込みベクトルの集合を θ で表している。Encoder でクエリ \mathbf{x} を $\mathbf{c}_x = enc_{\phi}(\mathbf{x})$ へと変換し、Decoder は \mathbf{c}_x と返答ユーザの個性表現ベクトル \mathbf{p}_{uid} を元に、隠れ層の系列を生成し単語への変換 g を用い生成単語系列へと変換する (図 1)。

$$P(y_m | y_{m-1}, y_{m-2}, \dots, y_1, x_l, x_{l-1}, \dots, x_1, \mathbf{p}_{uid}) \quad (2)$$

$$= P(y_m | y_{m-1}, y_{m-2}, \dots, y_1, \mathbf{c}_x, \mathbf{p}_{uid}) \quad (3)$$

$$= softmax(g(dec_{\psi}(\mathbf{h}_m, y_{m-1}, \mathbf{c}_x, \mathbf{p}_{uid}))). \quad (4)$$

Encoder, Decoder と共に返答ユーザ毎の個性表現ベクト

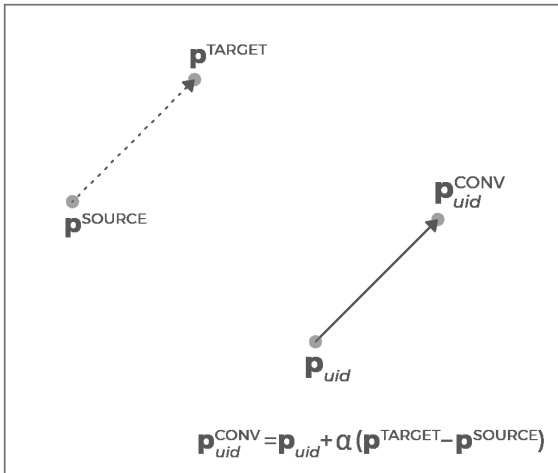


図 2 Transformation での個性表現ベクトル \mathbf{p}_{uid}^{CONV} の変換

ル \mathbf{p}_{uid} が学習されていくことにより、個性を反映した返答生成を行えるようになる。

3.2 個性表現ベクトルの変換 1: Transformation

表現ベクトルの加減算によって、表現ベクトルの対象となる事象の意味的な合成を実現できることが、幾つかの事例によって知られている。Mikolov ら [9] は単語の意味表現ベクトルにおいて、“Berlin” と “Germany” の表現ベクトルの差分が “国→首都” への変換ベクトルとなることを示した。また Larsen ら [4] や Upchurch ら [14] は、画像表現ベクトルの差分から “年配” や “笑顔” などの付加的な特徴のベクトルが得られることを示した。これらと同様に、Speaker Model の学習で獲得された返答の個性表現ベクトルを用いた返答文生成での個性追加を行う。特定の 2 ユーザの表現ベクトルの差分が特徴の変化を表すベクトルになっていることを期待して考えた手法が図 2 である。

具体的には、学習済みの返答ユーザの中から、追加対象の個性を持たないユーザと持つユーザを選び、それぞれの個性表現ベクトルを \mathbf{p}^{SOURCE} と \mathbf{p}^{TARGET} として、個性を追加反映するための方向を $\mathbf{p}^{TARGET} - \mathbf{p}^{SOURCE}$ で算出する。その差分ベクトルを個性の追加元の uid の個性表現ベクトル \mathbf{p}_{uid} に足すことで、個性の追加反映のための個性表現ベクトル \mathbf{p}_{uid}^{CONV} を得る。 α は、個性表現ベクトルを動かす大きさであり、個性が追加される傾向の強弱を調整するパラメータである。

$$\mathbf{p}_{uid}^{CONV} = \mathbf{p}_{uid} + \alpha (\mathbf{p}^{TARGET} - \mathbf{p}^{SOURCE}) \quad (5)$$

3.3 個性表現ベクトルの変換 2: Interpolation

元の返答ユーザ uid の個性表現ベクトル \mathbf{p}_{uid} を追加対象の個性を持つ返答ユーザ $TARGET$ の個性表現ベクトル \mathbf{p}^{TARGET} に近づけることで個性の追加反映を行う (図 3)。具体的には式 6 の 2 つの個性表現ベクトルの線形補完で、 \mathbf{p}_{uid}^{CONV} を得る。 $0 < \alpha < 1.0$ である。

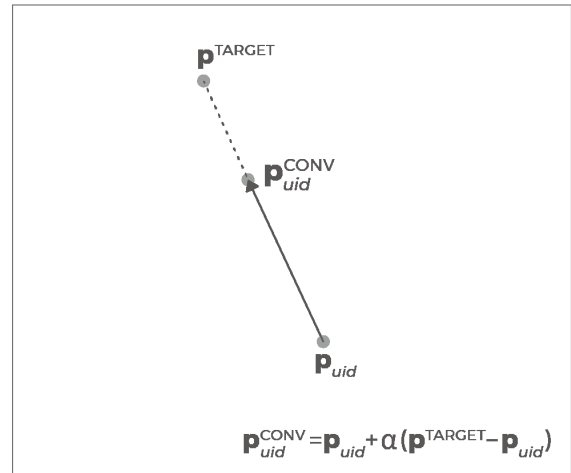


図 3 Interpolation での個性表現ベクトル \mathbf{p}_{uid}^{CONV} の変換

$$\mathbf{p}_{uid}^{CONV} = \mathbf{p}_{uid} + \alpha (\mathbf{p}^{TARGET} - \mathbf{p}_{uid}) \quad (6)$$

3.4 レスポンスの直接的書き換え: Translation

レスポンスの単語系列 $\mathbf{y} = (y_1, \dots, y_m)$ を、対象とする個性が追加反映されたレスポンスの単語系列 $\tilde{\mathbf{y}} = (\tilde{y}_1, \dots, \tilde{y}_m)$ に書き換える Encoder-Decoder を学習する。具体的には、式 7 の条件付き対数尤度を最大化する学習を行う。

$$\max_{\theta} \sum_{(\mathbf{y}, \tilde{\mathbf{y}}) \in D} \log p_{\theta}(\tilde{\mathbf{y}} | \mathbf{y}). \quad (7)$$

Encoder でレスポンスの単語系列 \mathbf{y} を $\mathbf{c}_{\mathbf{y}} = enc_{\phi}(\mathbf{y})$ へと変換し、Decoder は $\mathbf{c}_{\mathbf{y}}$ を元に個性が追加反映されたレスポンスを生成する。

$$P(\tilde{y}_n | \tilde{y}_{n-1}, \tilde{y}_{n-2}, \dots, \tilde{y}_1, y_m, y_{m-1}, \dots, y_1) \quad (8)$$

$$= P(\tilde{y}_n | \tilde{y}_{n-1}, \tilde{y}_{n-2}, \dots, \tilde{y}_1, \mathbf{c}_{\mathbf{y}}) \quad (9)$$

$$= \text{softmax} \left(g \left(dec_{\phi}(\mathbf{h}_n, \tilde{y}_{n-1}, \mathbf{c}_{\mathbf{y}}) \right) \right). \quad (10)$$

学習時には、全てを強引に書き換えるモデルにならないように、まず変換前のレスポンスの単語系列 $\mathbf{y} \rightarrow \mathbf{y}$ を復元する Encoder-Decoder を学習する。次いで、個性の追加反映の書き換え $\mathbf{y} \rightarrow \tilde{\mathbf{y}}$ の訓練データを使った追加学習でパラメタ全体を更新する。

4. 評価実験

本節では、追加対象の個性の反映がどの程度可能であるかを定量的に評価した結果を述べる。それぞれの個性の追加反映の手法で、個性対象のユーザの個性表現ベクトルを用いて生成した返答から基本的な内容を変えずに、追加対象の個性が反映できた割合を主観評価した。

4.1 データセットの作成

学習・評価対象とするデータとして、株式会社ディー・エ

ス・エーが提供するゲームプラットフォーム「Mobage」上でのオープンチャットコーナー「みんなとチャットβ」*1におけるユーザの対話ログを用いた。当サービスは多人数チャットであり、各発話に対応する返答が機械的に算出することができなかったため、各発話に対して返答関係にある発話をアノテーションし、Mobage データとした。

また、3種類の個性を用いて追加反映を評価するため、4種のデータ (*Neutral*, *Kaomoji*, *Kansai*, *Aniki*) を人手で作成した。*Neutral* は、Mobage の中からクエリを 6,200 件抽出し、方言や顔文字など個性を自然な範囲で除外した状態で返答文を作成したデータであり、3.2 で示した p^{SOURCE} に用いる。*Kaomoji*, *Kansai*, *Aniki* は *Neutral* と同じ 6,200 件のクエリに対し、内容は変えずに表現のみそれぞれの個性 (顔文字など装飾文字を多用して返答、関西弁で返答、熱血漢が感じられるよう返答) を反映するよう作成したデータである。これら 3 種類のデータは、*Kaomoji* は比較的規則的で文そのものを変更しない程度の変更、*Kansai* は軽微な単語の変更や語末の表現の変更、*Aniki* は表現全体が大きく変わる可能性のある変更に対してそれぞれ対応できるか確認することが目的となっている。各データセットの内容を表 1、具体的な事例を表 2 にそれぞれ示す。

4.2 モデルの学習

ベースモデルと *Translation* とともに、Encoder と Decoder には 3 層の多層 LSTM モデルを採用し、単語ベクトルと隠れ層の次元数を 1,000 次元、dropout rate を 0.2、バッチサイズ 128 で、ミニバッチ学習を行った。

語彙はデータセット内の頻度上位 20,000 単語を用い、形態素解析には MeCab[18] を利用した。各パラメータの学習は Adam[3] を利用し、Adam の α の値については、ベースモデルでは 0.001、*Translation* では 0.0001 を用いた。

Neutral, *Kaomoji*, *Kansai*, *Aniki* の 4 種類を 4 人の仮想ユーザのデータとみなし、Mobage データの約 23 万個の (クエリ、返答ユーザ ID、レスポンス) に追加し、訓練、開発、テストで 8:1:1 にランダムに分割して、訓練データを用いて 50epoch までモデルを学習し、開発データの BLEU と生成例を見て主観的に個性が反映されていると考えられたモデルを選択した。

4.3 評価用データの作成

開発データで選んだベースモデルを用いて、テストデータに対してレスポンスを生成し、それらをランダムシャッフルして、それぞれの個性に対して、追加反映できると考えられる (クエリ、返答ユーザ ID、レスポンス) を 100 例ずつ人手で選出した。これにより、評価用のデータは、全

て追加対象の個性を付与できると考えられるデータとなっている。

4.4 個性表現ベクトル変換のパラメタ設定

開発用データでいくつかのパラメタを設定して生成例を作成し、その結果を見ながら経験的に設定した。*Transformation* では式 5 の α は 3 種類とも 2.0 に設定した。*Interpolation* では式 6 の α の値を *Kaomoji*, *Kansai*, *Aniki* に対してそれぞれ 0.75, 0.75, 0.5 に設定した。*Transformation* の p^{SOURCE} には *Neutral* の個性表現ベクトルを用い、*Transofromation*, *Interpolation* 共通で、 p^{TARGET} にはそれぞれの個性に対応する仮想ユーザの個性表現ベクトルを用いた。

4.5 主観評価の結果

それぞれの評価対象の個性について、3.2, 3.3, 3.4 で提案した 3 つの手法の出力を合わせたそれぞれ 100 個ずつ 300 個をシャッフルして、著者らを含まない被験者 3 名に以下の観点で独立に評価してもらい、最終的には 2 名以上の一致を評価結果とした。

- 返答内容が基本的には変わっていないか (内容保持)
- 追加対象とした個性を付与できているか (個性付与)

表 3 に 3 名の被験者の評価の Fleiss' Kappa の値を示す。

主観評価の結果を表 4, 表 5, 表 6 に示す。太字は、個性毎に最も評価が高かった手法を示している。総合評価は表 4 であるが、*Kaomoji*, *Kansai*, *Aniki* 全ての個性において *Translation* が一番良い結果となった。表 5 に示す通り、変換後に内容が保持される割合が *Translation* のみ大きいいため、両方の条件を求める総合評価に影響した結果となった。これは、*Translation* のみ、元の内容を残しながら個性を追加反映するために直接的な学習を行なっているため、妥当な結果と言えよう。一方、表 6 に示す通り、個性付与の割合は拮抗する結果となった。

5. 生成例の分析

前節では、定量的な主観評価について述べたが、本節では、生成されたレスポンスの特徴の分析を通して、今後の方向性を議論する。

5.1 モデル・追加個性ごとに見た生成例の特徴

表 7 に、同じクエリに対する手法ごとのレスポンスの違いの例を示す。総じて、*Translation* は、単語列の一部のみ書き換えて個性を追加反映している例が多かった。これは、単語列を復元できるモデルを事前学習していることと、追加学習のための訓練データが基本的な内容を変えないで作成したという制約から、一部のみが書き換わっている例が多いことから、想定される動作と言える。また、*Translation* は個性を追加反映させるように無理に書き換

*1 http://sp.mbga.jp/_chat_app

表 1 実験用に作成したデータの概要

	対話ペア数	ユーザ数	概要
<i>Mobage</i>	226,081	1379	Mobage みんなとチャットβ上でユーザ間で行われた会話データを基に作成したデータセット
<i>Neutral</i>	6,200	1	方言や顔文字など個性を自然な範囲で除外したデータセット
<i>Kaomoji</i>	6,200	1	顔文字や記号など装飾文字を自然な範囲で付与するようにしたデータセット
<i>Kansai</i>	6,200	1	関西弁で返答するようにしたデータセット
<i>Aniki</i>	6,200	1	熱血感が感じられるように返答するようにしたデータセット

表 2 各データセット中の対話ペア例

発話	うそだよ
<i>Neutral</i> の返答	嘘つくのダメ!
<i>Kaomoji</i> の返答	嘘つくのダメ!
<i>Kansai</i> の返答	嘘つくのアカン!
<i>Aniki</i> の返答	嘘つきはスポーツマンシップに反する!
発話	お騒がせ申し訳ない
<i>Neutral</i> の返答	いいのよ
<i>Kaomoji</i> の返答	いいのよ (*~*)
<i>Kansai</i> の返答	ええんやで
<i>Aniki</i> の返答	いいんだぞ

表 3 3名の主観評価の Fleiss' Kappa

	<i>Kaomoji</i>	<i>Kansai</i>	<i>Aniki</i>
内容保持	0.89	0.79	0.80
個性付与	0.98	0.83	0.46

表 4 内容が保持され、かつ個性付与された割合

手法	<i>Kaomoji</i>	<i>Kansai</i>	<i>Aniki</i>
Translation	59%	45%	22%
Transformation	33%	21%	11%
Interpolation	18%	17%	4%

表 5 内容が保持された割合

手法	<i>Kaomoji</i>	<i>Kansai</i>	<i>Aniki</i>
Translation	74%	79%	96%
Transformation	41%	35%	26%
Interpolation	30%	18%	16%

表 6 個性付与された割合

手法	<i>Kaomoji</i>	<i>Kansai</i>	<i>Aniki</i>
Translation	77%	55%	25%
Transformation	76%	50%	30%
Interpolation	59%	57%	20%

える傾向がみられた。*Kansai*と*Aniki*では、100例中で、それぞれ13個、15個、言葉として不自然になっていた。例えば、*Kansai*では「なんだよ」を「なんやで」としたり、「そうなんだ」を「そおなんやんだ」と変換されたり、*Aniki*では、「うん」が「うむじゃねーか!」に変換されていた。この手の言語としての崩れは*Transformation*で5個以下、*Interpolation*では*Kansai*で1個のみと、ほぼ見られなかった。

一方、*Transformation*は、内容を保持しつつ全体を書き換えた上での個性の反映ができていた例が見られた。例え

ば、*Kaomoji*では、顔文字自体で元のレスポンスの内容を表している。また、*Kansai*でクエリにあった「沖縄」がレスポンスに入っている。*Aniki*では、元の個性である関西弁を「うち」として残しつつ個性が反映されたレスポンスを生成できている。これは、返答生成モデルに対して直接的に個性表現ベクトルを作用させているからこそその特性と考えられる。

*Interpolation*の生成例からは特筆した傾向は掴めなかったが、今回、個性を反映させるために、追加対象の個性を持つ返答ユーザの影響を強めた結果、内容保持がされない例が多かった。

5.2 個性表現ベクトルの変換における係数変化に応じた生成例の特徴

表 8 は、*Transformation*と*Interpolation*で、個性を反映するための係数 α を変化させた生成例である。どちらの手法も $\alpha = 0.0$ が元の返答ユーザと一致する。また、*Interpolation*では、 α が1.0の場合に、追加対象の個性を持つ返答ユーザと一致する。

全体として、内容の保持がされているとは言い難いが、 α の値をあげることで、個性付与の程度が調整できている傾向は伺える。例えば、*Aniki*では α の値を上げていくにつれてまず記号「!」が付き、次第に*Aniki*の個性である真面目でストイックな内容が強まっている。また、最適な α は事例や個性により異なっていることから、固定の α を使うのではなく、何かしらの基準から相対的に決定することで改善の余地があると考えられる。これは今後の課題である。

*Transformation*では、 $\alpha \geq 3.0$ から、言語として崩れ出す傾向があった。これは、元の返答ユーザに対してベクトルを足した結果、学習済みの個性表現ベクトルの分布から外れたためと考えられる。この仮説のための分析自体が今後の課題であるが、返答生成のDecoderを壊さないように個性表現ベクトルを動せる範囲を検討する一助になると考えている。

*Interpolation*では、 α の0.5から個性が付与されはじめ、0.75まで行くと1.0(追加対象の個性を持つ返答ユーザと一致)とほぼ同じ生成例となっていた。今回の実験では、個性が付与される割合を重視して α に設定したが、その結果、想定よりも追加対象の個性を持つ返答ユーザと同じ生

表 7 生成されるレスポンスの違いの例

追加個性	Kaomoji	Kansai	Aniki
クエリ	はじめましてよろしくね	沖縄は年に数回かな	まあまあかかるね(´・д・`)笑 関東住み？
Speaker Model	よろしくー	いいねー☆☆☆☆	そやで(笑)
Interpolation	よろしくです！	ほお～	ちがうよ～
Transformation	こちらこそ30(´・ω・`)矧！	沖縄ええなあ	うちは東京だぜ！
Translation	よろしくー	ええねー☆☆☆☆	そうぞぞ！
クエリ	でわでわトコトコ(((((*´ω´*))	でも、お部屋でエアコンガンガンやろ？	今から行ってくるよ、仕事(´▽´)
Speaker Model	はーい	そだよ～(n´▽´)n	頑張れ頑張れ
Interpolation	ガリガリ☆*. (*´▽´*)ノ	そおなん？	お疲れ様です
Transformation	いー！	そやで	おつかれだな！
Translation	はーい(*´ω´*)ノ	せやで～(n´▽´)n	頑張れ頑張れ！

成結果となる場合が多くなってしまったと考えられる。

6. おわりに

本研究では、雑談の返答生成において、キャラクタ設定としての個性を追加反映できる雑談対話エージェントを提案し、評価した結果を報告した。

評価実験では、それぞれの手法の特性を評価するため、生成されたレスポンスに追加対象の個性が付与されているか否か、元の内容を維持しているか否かを観点に主観評価を行った。生成例の分析を通して、*Translation* は元のレスポンスを単語単位で保持しやすい特性があり、*Transformation* は内容を保持しつつも全体的に変換して個性を反映できる特性が見て取れた。

今回3つの個性について実験、評価を行ったが、今後より多様な個性に対して適用対象を広げていく。また今回実験でわかった手法ごとの特性を活かし、3つの手法を状況に応じて組み合わせることで、より適切な個性の追加反映を行うことを検討していく。

参考文献

[1] Higashinaka, R., Kobayashi, N., Hirano, T., Miyazaki, C., Meguro, T., Makino, T. and Matsuo, Y.: Syntactic filtering and content-based retrieval of Twitter sentences for the generation of system utterances in dialogue systems, *Situated Dialog in Speech-Based Human-Computer Interaction*, Springer, pp. 15–26 (2016).

[2] Hu, Z., Yang, Z., Liang, X., Salakhutdinov, R. and Xing, E. P.: Controllable Text Generation, *CoRR*, Vol. abs/1703.00955 (online), available from <http://arxiv.org/abs/1703.00955> (2017).

[3] Kingma, D. and Ba, J.: Adam: A method for stochastic optimization, *In The International Conference on Learning Representations (ICLR)* (2015).

[4] Larsen, A. B. L., Sønderby, S. K., Larochelle, H. and Winther, O.: Autoencoding Beyond Pixels Using a Learned Similarity Metric, *Proceedings of the 33rd International Conference on International Conference on Machine Learning (ICML)*, Vol. 48, pp. 1558–1566 (2016).

[5] Li, J., Galley, M., Brockett, C., Gao, J. and Dolan, B.: A Diversity-Promoting Objective Function for Neural Conversation Models, *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language*

Technologies, pp. 110–119 (2016).

[6] Li, J., Galley, M., Brockett, C., Spithourakis, G., Gao, J. and Dolan, B.: A Persona-Based Neural Conversation Model, *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 994–1003 (2016).

[7] Mairesse, F. and Walker, M. A.: PERSONAGE: Personality Generation for Dialogue, *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*, pp. 496–503 (2007).

[8] Mikolov, T., Karafiát, M., Burget, L., Cernocký, J. and Khudanpur, S.: Recurrent neural network based language model., *Interspeech*, Vol. 2, p. 3 (2010).

[9] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S. and Dean, J.: Distributed representations of words and phrases and their compositionality, *Advances in neural information processing systems*, pp. 3111–3119 (2013).

[10] Mizukami, M., Neubig, G., Sakti, S., Toda, T. and Nakamura, S.: Linguistic individuality transformation for spoken language, *Natural Language Dialog Systems and Intelligent Assistants*, pp. 129–143 (2015).

[11] Ritter, A., Cherry, C. and Dolan, W. B.: Data-driven Response Generation in Social Media, *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP '11)*, pp. 583–593 (2011).

[12] Shibata, M., Nishiguchi, T. and Tomiura, Y.: Dialog System for Open-Ended Conversation Using Web Documents, *Informatika (Slovenia)*, Vol. 33, No. 3, pp. 277–284 (2009).

[13] Sutskever, I., Vinyals, O. and Le, Q. V.: Sequence to sequence learning with neural networks, *Advances in neural information processing systems*, pp. 3104–3112 (2014).

[14] Upchurch, P., Gardner, J., Pleiss, G., Pless, R., Snavely, N., Bala, K. and Weinberger, K.: Deep Feature Interpolation for Image Content Changes, *Computer Vision and Pattern Recognition (CVPR)* (2017).

[15] Vinyals, O. and Le, Q. V.: A Neural Conversational Model, *CoRR*, Vol. abs/1506.05869 (2015).

[16] Xing, C., Wu, W., Wu, Y., Liu, J., Huang, Y., Zhou, M. and Ma, W.: Topic Aware Neural Response Generation, *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, pp. 3351–3357 (2017).

[17] Zhou, G., Luo, P., Cao, R., Lin, F., Chen, B. and He, Q.: [Mechanism-Aware Neural Machine for Dialogue Response Generation, *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, pp. 3400–3407 (2017).

[18] 工藤拓, 山本薫, 松本裕治: Conditional Random Fields を用いた日本語形態素解析, 情報処理学会自然言語処理研究会 SIGNL-161, Vol. 47, pp. 89–96 (2004).

表 8 係数ごとのレスポンス生成例の変化

	Kaomoji		Kansai		Aniki	
	Transformation	Interpolation	Transformation	Interpolation	Transformation	Interpolation
クエリ	そうなんですかあ。私も仕事ですけど、今夜動中(笑)		今日は沖繩の姉が来てるので食事会です		お仕事行きたくない	
$\alpha=0.0$	そうなんですね。	そうなんですね。	え、そうなんだ!	え、そうなんだ!	大丈夫?	大丈夫?
$\alpha=0.25$	そうなんですね。	そうなんだ。	えー!?	そうなんですね!	仕事頑張ってるね	仕事頑張ってるね
$\alpha=0.50$	そうなんです。	そうなんだ!	沖繩いいなあ	沖繩いいなあ	いつてらっしゃい!	がんばれよ! 負けんな! 熱血!
$\alpha=0.75$	お疲れ様です。	仕事かあ(´ω´)おつかれさー	沖繩ええなあ	沖繩ええなあ〜!	お疲れ様です!	がんばれよ! 負けんな! 熱血!
$\alpha=1.0$	お疲れ様です。	仕事かあ(´ω´)おつかれさー	沖繩ええなあ	沖繩ええなあ〜!	頑張ってくれ!	がんばれよ! 負けんな! 熱血!
$\alpha=2.0$	お疲れ様です。	-	沖繩ええなあ	-	仕事頑張れ!	-
$\alpha=3.0$	夜勤お疲れ様(^-^)	-	沖繩ええなあ	-	仕事頑張れ!	-
$\alpha=10.0$	仕事!	-	沖繩?、、、、詳しいー	-	仕事頑張れ!	-
クエリ	邪魔じゃないよ(笑)		そっか(´-´*)w仕事大変なのかなw		あとは夜仲間と絡む〜www	
$\alpha=0.0$	(笑)	(笑)	そうだよ。	そうだよ。	そうなん?	そうなん?
$\alpha=0.25$	(笑)	(笑)	そうだよ。	そうだよ	そうなん?	そうなん?
$\alpha=0.50$	うん(笑)	ω´) チラッ	そうだよ。	大変だよね〜	そうなんや!	おいっ!
$\alpha=0.75$	そか(笑)	(´0´)おーおー	そうだよ。	大変やろうね〜	そうなんや!	おいっ!
$\alpha=1.0$	分かってるよ♪	(´0´)おーおー	大変だよね。	せやで〜	そうなんや!	おいっ!
$\alpha=2.0$	分かってるよ(´艸´)	-	仕事やで〜	-	何を!?	-
$\alpha=3.0$	分かっ ((´▽´)	-	そやで〜	-	そっな!	-
$\alpha=10.0$	(´▽´)(^´)	-	うん大変でで	-	何!	-

- [19] 宮崎千明, 平野 徹, 東中竜一郎: 発話にキャラクター性を与えるための文末表現の変換, 人工知能学会言語・音声理解と対話処理研究会, Vol. 68, pp. 41-46 (2013).
- [20] 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ー性変換, 人工知能学会論文誌, Vol. 31, No. 1 (2016).
- [21] 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).
- [22] 沈 睿, 菊池英明, 太田克己, 三田村健: 音声生成を前提としたテキストレベルでのキャラクター付与, 情報処理学会論文誌, Vol. 53, No. 4, pp. 1269-1276 (2012).