

# 説得対話システムにおける感情表現を反映させた 応答生成モデルの構築

浅井 沙良<sup>1,a)</sup> 品川 政太郎<sup>1,3</sup> 吉野 幸一郎<sup>1,2,3</sup> サクリアニ サクティ<sup>1,3</sup> 中村 哲<sup>1,3</sup>

**概要：**感情表現が説得成功率の向上に寄与することが知られており、感情表現を用いた説得対話システムの研究が行われてきた。こうした説得を目的とするシステムは、説得のために最適化された感情状態と文脈に応じた自然な発話を行う必要がある。条件付き文生成の研究では、条件ラベルを言語モデル・デコーダに入力することが行われてきたが、与えた条件が必ずしも反映されないという問題があった。本研究では、文生成を行う際のデコーダで入力された感情ラベルを予測することで、指定された感情状態を反映させた応答文を生成する手法を提案した。

## 1. はじめに

説得対話とは対話を通じて説得者が意図する状態に相手を誘導するものである。例えば、商品の購入を促す販売やサークルへの勧誘、日常で家事を促す場面などは説得対話として解釈することが出来る。近年、Captorogy[1]という分野において、説得などユーザの行動変容を促すことを目的とした研究が行われている。説得には対話内容や伝え方、相手との関係性、両者の感情状態など様々な要素が影響すると考えられている。本研究では、その中でも感情表現に着目して説得成功率を向上させるシステムの構築を目指す。

人間が説得を行う場合、説得者の感情表現が説得の成功率に影響を与えることが認められている [2][3]。例えば、説得者がポジティブな感情を持っている場合に、相手に良い印象を与え説得の成功率が向上することが知られている [4]。また、対話の状況によってはネガティブな感情表現が効果的である場合があることも認められている。例えば、相手がどの選択肢にも特段の魅力を感じていない場合に、説得者が「怒り」の感情を表現することで、主張を強化し相手の譲歩を促す効果があるとされている [5]。さらに、システムによる説得においても、システムが感情表現を用いることで説得の成功率が向上する可能性が示唆されている [6]。この場合、システムは説得の文脈に応じて適切に感情が反映された応答を行う必要がある。つまり、対話

中のシステムの感情状態の制御と、対話の流れに沿った内容を保持した上で感情表現を反映した応答の制御を、それぞれ行う必要がある。

本研究では、説得対話システムの説得成功率の向上を目指し、入力としてユーザ発話文だけでなくシステムに表現させたい感情も同時に与えることで、入力感情を反映させた応答文を生成する。この際、より生成文に感情表現を反映させるための工夫として、表現させたい感情ラベルをデコーダへの入力として与える従来のモデルに加え、デコーダで与えた感情ラベルを予測するモデルを構築する。実験では、構築したモデルに対して、感情の反映度、対話の自然性、説得への効果の観点で評価を行った。

## 2. 関連研究

### 2.1 説得対話システムの研究

説得対話システムの研究は、主に対話制御によって説得性を上げることに焦点がおかれている [8][9][10]。Geogilara [8] はシステムの持つ目的に応じて設定した報酬を強化学習によって最大化した。また、Hiraokaら [9] はシステムの説得成功だけでなく、ユーザの満足度も向上するような対話戦略の探索を行う手法を提案している。これらの研究では、システムは論理的な説得を行うことにとどまっておらず、感情状態の考慮は行われていない。

石川ら [11] は説得対話システムが感情表現を用いる方法を提案した。この研究では、スコア計算はコーパスから算出した遷移確率に従って行われており、一発話前のユーザ発話内容とシステムの感情状態に沿った応答文選択が行われている。しかし、システムの応答生成は単純な応答選択にとどまっている。

<sup>1</sup> 奈良先端科学技術大学院大学  
Nara Institute of Science and Technology  
<sup>2</sup> 科学技術振興機構 さきがけ  
<sup>3</sup> 理化学研究所 革新知能統合研究センター AIP  
<sup>a)</sup> asai.sara.ar6@is.nasit.jp

こうした応答選択に基づくシステムの課題として、対話の文脈や条件によって選択可能な応答候補数が極端に減少し、応答が不自然になることが挙げられる。また、感情を用いた説得対話システムにおいて説得成功率を向上させるためには、対話の流れに沿った感情遷移を行わせるだけでなく、説得に効果的な感情状態に変化させることが望ましい。そこで本研究では、自然な応答内容を保持したまま感情表現の制御を可能にする応答生成モデルの構築を目指す。

## 2.2 感情表現を含む応答生成モデルの研究

近年 Neural Conversation Model (NCM) 研究の進展から、雑談システムで感情表現を応答生成に組み込む研究が行われている。Ghosh[12] は感情表現を行う言語モデルを提案しており、感情極性を持つ単語の割合を調整することで発話文中の感情表現の制御を可能にしている。対話モデルでは、Zhou ら [13] は内部メモリにおいて動的な感情状態の変化を応答文に反映させ、また外部メモリによる表層的な単語置き換えにより、発話に対する応答文を感情ごとに変化させるモデルを提案している。また Li ら [14] は、強化学習を用いて Zhou らの応答生成モデルの最適化を行っている。

こうしたモデルの多くは、大規模な学習データを必要とする。しかし、説得対話はそのドメインに応じて発話内容が異なり、ドメインごとに大規模データを用意することは難しい。また、多くのモデルがニューラルネットワークの最適化対象を単語や文においているため、ラベル情報よりも言語モデル上の自然さが優先されがちになる。そこで本研究では、他ドメインでの大規模データによる事前学習を行うと共に、入力された感情ラベルをデコーダで再予測することによって、入力された感情ラベルを反映した説得発話の生成を目指す。

## 3. 応答生成モデルの構築

### 3.1 感情表現を反映した説得発話生成

本研究で構築するモデルは、説得成功率を最大化するように決定された感情ラベル ( $e$ ) と対話の文脈 (ユーザ発話) を用いて生成を行う。この生成結果は、以下の条件を満たすような応答生成モデルが期待される。

- (1) 目的の説得をしていること (e.g. 掃除を促す, 健康的な食事を促す, 就寝を促す)
- (2) 応答として自然であること
- (3) 表現させたい感情を反映させた応答を行うこと

### 3.2 感情状態の定義

本研究で構築する応答生成モデルで用いる感情状態を、図 3.2 に示す。Russell の円環モデル [7] を用いた 4 つの感情を用いる。Russel の円環モデルは、縦軸を活発-不活

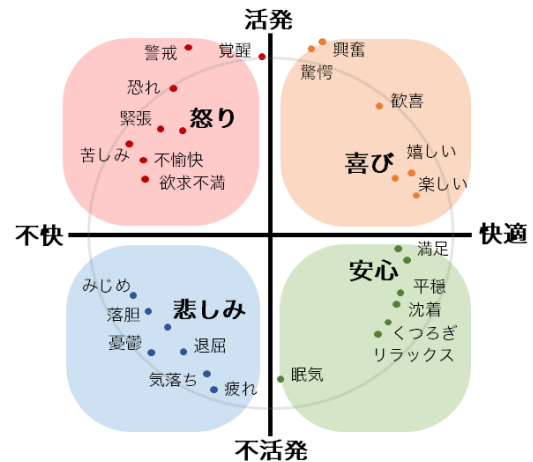


図 1 Russell の円環モデル. 二軸で区切られたそれぞれの空間を一つの感情としてまとめて扱う。中心部分ほどの感情も表現されていない「平静」とする。

発、横軸を快適-不快と定義されており、様々な感情が二次元平面上で定義されている。その二軸で区切られた各象限を一つの感情として扱い、4 つの感情は「怒り」「悲しみ」「喜び」「安心」として与える。これらのいずれにも属さない中央の領域を「平静」とする。石川らは [11] はこの感情状態の定義に従った感情ラベルを付与した対話コーパスを収集している。本研究では石川らのデータセットを用いて学習を行うため、同様の感情を採用した。

## 4. 感情ラベルを同時予測する発話文生成

感情ラベルなどのクラスラベルを応答に反映させたい場合、最もよく使われる手法はデコーダの各ステップにクラスラベルを入力するモデルである。しかし、デコーダへの入力としてクラスラベルを用いるだけでは、出力がクラスラベルを反映した状態で学習されるという保証がない。これに対し本研究では、デコード終了時にデコーダの各ステップで入力された感情のクラスラベル  $e$  を予測し、学習時にこの損失を逆伝播させることでこの問題の解決を試みる (図 2)。本研究が提案する感情ラベルを用いた応答制御の例を表 1 に示す。本書では、まずデコードの各時点でラベル情報を与えるベースラインについて説明し、提案するデコーダにおいてラベル予測も行うモデルの説明を行う。

### 4.1 ベースラインモデル

まず、ベースラインとなる感情状態を考慮した応答生成モデルには、再起型ニューラルネットワーク (RNN) に基づくエンコーダデコーダモデル [15] を用いる。RNN の実装には、Gated Recurrent Unit (GRU) [18] を用いる。GRU を用いたエンコーダでは、発話中の単語系列  $X = (x_1, x_2, x_3, \dots, x_m)$  を入力したとき、各ステップ  $t$  に

表 1 感情を反映させた応答生成モデルの活用例

話者	入力感情	発話文
ユーザ	-	掃除は嫌だし片付けなくていいかな
システム	喜び	片付け楽しいよ!
ユーザ	-	別にいいでしょう
システム	悲しみ	何で片付けてくれないんですか....
ユーザ	-	片付けたくないから
システム	怒り	汚い部屋にいて気分悪いし片付けてよ!!
ユーザ	-	仕方ないな...
システム	安心	ありがとう, 分かってくれて嬉しい

おける隠れ層  $h_t^{enc}$  を

$$h_t^{enc} = \text{GRU}(h_{t-1}^{enc}, x_t) \quad (1)$$

によって更新する. ここで GRU は GRU による重みの更新を表す関数とする. 今回は  $x_t$  として発話中の各単語を分散表現によりベクトル化したものを用いる.

一般的な GRU のデコーダでは, 応答の単語系列  $Y = (y_1, y_2, y_3, \dots, y_n)$  を予測するが, エンコーダデコーダモデルでは各ステップ  $t$  の入力として, 直前の予測単語  $y_{t-1}$  を用いる. つまり, ステップ  $t$  におけるデコーダの隠れ層は,

$$h_t^{dec} = \text{GRU}(h_{t-1}^{dec}, y_{t-1}) \quad (2)$$

によって更新する. この隠れ層に対して,

$$y_t = \text{softmax}(W_{hy}h_t^{dec} + b_y) \quad (3)$$

を適用して出力単語  $y_t$  を求める. ここで  $b_y$  はバイアス項であり, softmax はソフトマックス関数である.

ベースラインモデルでは, このデコーダ中の隠れ層の更新において, 感情ラベルをベクトル化したもの  $e_v$  を入力する. GRU の各ステップにおける入力  $y_{t-1}$  に対して  $e_v$  を連結し,

$$h_t^{dec} = \text{GRU}(h_{t-1}^{dec}, y_{t-1} \oplus e_v) \quad (4)$$

として用いる. ただし,  $\oplus$  はベクトルの連結を表す.

感情ラベルには, 平静, 怒り, 悲しみ, 喜び, 安心のクラスを用い. 学習時にはシステム発話にアノテーションされた感情ラベルを用いる. 事前学習時は感情ラベルを NONE に設定し, 学習に感情ラベルが影響しないよう設定する. 今回用いた対話コーパスは 3 人以上のアノテータが感情ラベルを付与しているが, これが 3 つとも一致しない場合も同様に NONE として学習する.

#### 4.2 提案モデル

ベースラインモデルでは, 感情ラベルが生成に反映されるという保証がない. そこで提案モデルでは, デコード時の予測単語列  $Y = (y_1, y_2, \dots, y_t)$  の末尾に感情ラベルに対応する単語を付け加える. 例えば図 2 の例では, デコーダに与える感情ラベル  $e$  が怒りの時は, 怒りの感情ラベル

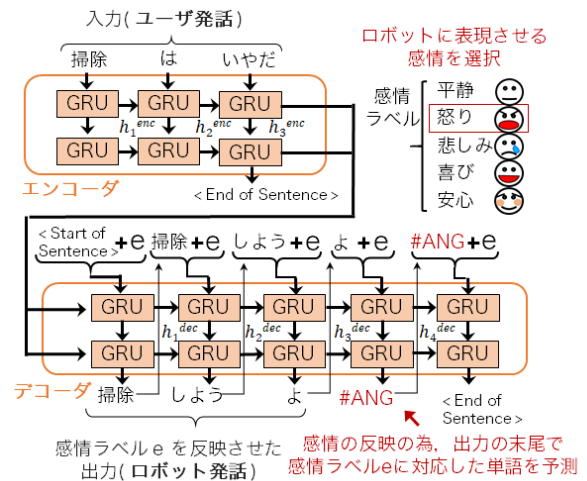


図 2 システムの概要図

に対応する単語である #ANG を用いて,  $y_{n+1} = \#ANG$  のように付け加える. デコード時に感情ラベルの情報を毎度与えるだけでなく, 出力時に与えた感情ラベルを予測し単語として出力する処理を加えることで, より感情を反映した出力文が生成されることを期待する.

#### 4.3 事前学習

モデルの学習は, 3 段階に分けて行う. まず大規模な対話データを用いて事前学習を行い, その後, 感情ラベルが付与された複数の説得目的の対話データで事前学習し, 最後に感情ラベルが付与された説得対話データの 1 つの目的の対話における応答生成モデルを学習する.

### 5. 実験設定

#### 5.1 データ

モデルの学習には, Twitter のリプライのデータセットと石川らによって収集された感情ラベル付き説得対話コーパス [11] を用いる. Twitter のリプライデータは応答生成モデルを事前学習する際に用い, その後, 石川らのデータを用いて学習を行った. それぞれのコーパスの発話数を表 2 で示す. Twitter のリプライデータは評価データとテストデータを 5000 発話ペアずつ抽出し, それ以外を学習データとした. また, 感情ラベル付き説得対話コーパスは評価データとテストデータを約 200 発話ペアずつ (20 対話ずつ) 抽出し, それ以外を学習データとした.

感情ラベル付き説得対話データのコーパスは, クラウドソーシング上でクラウドワーカーによって作成されたシナリオが収録されている. 日常的に生活を共にする生活指導ロボットとユーザとの対話を想定した 5 つの場面「掃除をすること」「食べ残しをしないこと」「早く寝ること」「ゲームを止めること」「適度な運動をすること」におけるシナリオが収集されている. 各対話は約 10 発話ペア (ユーザ発話一ロボット発話) を含む. 感情ラベルはすべての発話ご

表 2 学習データの発話ペア数

Twitter リプライ	感情ラベル付き説得対話コーパス					
	掃除	食事	就寝	ゲーム	運動	合計
2,823,875	2,292	2,326	2,180	2,155	2,219	11,172

表 3 ドメインごとのロボット発話に対応する感情ラベルの個数

	掃除	食事	就寝	ゲーム	運動
平静	525	<b>716</b>	<b>829</b>	<b>682</b>	<b>813</b>
怒り	550	422	481	602	326
悲しみ	<b>598</b>	504	388	436	396
喜び	246	380	163	143	380
安心	33	56	107	46	56
NONE	340	248	212	246	248

とに 3 人のアノテータによってアノテーションがされている。今回の実験はロボット側の感情表現を行うことが目的であるため、ロボット発話の感情ラベルのみを用いる。

コーパスに含まれる感情ラベルの割合にはばらつきがあるため、学習の際に最も多い感情ラベルの発話数に合わせ、他の感情ラベルの発話をオーバーサンプリングして学習を行った。オーバーサンプリングを行う前のコーパス内の発話ペアにおける感情ラベルの個数をそれぞれ表 3 に示す。テストには石川らのデータセットから分割したテストデータを用い、直前のユーザ発話及びロボット (システム) の感情として付与されているラベルをモデルへの入力とした。

## 5.2 モデルの設定

応答生成モデルの学習には、エンコーダとデコーダともに、隠れ層 256 次元の GRU 2 層、バッチサイズ 64、Dropout 率 0.1、teacher forcing 率 1.0、Optimizer を Adam[17] とし、Gradient Clipping 50、学習率  $3e^{-6}$  (大規模なデータによる事前学習時は学習率  $3e^{-4}$ )、損失関数を Inverse Token Frequency (ITF) loss [19] とした。トークン分割には sentencepiece[16] を用い、語彙数 32,000 とした。単語分散表現の次元数は 256、デコーダにおいて単語分散表現と連結する感情ラベルの分散表現の次元数を 64 とした。応答生成時はビームサーチを用い、ビーム幅 20、探索の際 Repetitive Suppression[19]、length normalization[20] を用いた。

## 6. 評価

### 6.1 自動評価によるモデルの比較

ベースラインモデルと提案モデルの比較を行うため、まず、参照文を用いた自動評価を行う。自動評価では、まず応答の自然性を評価するため、テストデータに存在する参照文に対する BLEU[21] を計測した (表 4)。なお、BLEU の算出は、出力文を連結し MeCab<sup>\*1</sup> で再分割した上で行っ

<sup>\*1</sup> <https://taku910.github.io/mecab/>

表 4 応答の自然性の自動評価

	BLEU(正解文, 出力文)
感情ラベル不使用	<b>0.944</b>
ベースライン	0.772
提案モデル	<b>0.897</b>

表 5 感情の反映度合いの自動評価

	違う文の出力割合
ベースライン	0.850
提案モデル	<b>0.932</b>

た。表 4 より、ベースラインモデルでは感情ラベルの入力をしない場合と比較して自然性が低下してしまう。これに対して提案モデルは、感情ラベルの入力を行わない場合と同等の BLEU を達成している。このことから提案モデルのほうが、より自然な応答を出力できていると考えられる。

また、感情の反映度を調査するため、「平静」ラベルを与えて出力した場合と他の感情ラベルを与えて出力した場合で異なる文章を出力している割合を算出した。(表 5)。ベースラインモデルよりも提案モデルのほうが、違う文章を出力している割合が 8.2% 高いことが見て取れる。このことから、提案モデルのほうが感情表現を反映できている可能性が高いと考えられる。

### 6.2 主観評価

前述した自動評価だけでは、実際の生成応答の自然性、感情の反映度合いなどを確認することが難しい。そこで、感情ラベルの反映度合い、説得としての有効性、応答としての自然性の 3 つの観点から人手による主観評価を行った。主観評価は 54 名 (男性 19 名、女性 35 名) を対象に行った。

#### 6.2.1 感情ラベルの反映度合いを測る質問

感情ラベルの反映度合いを測る質問は、入力発話を提示したうえで、入力感情に「平静」を入れた場合と目標とする感情ラベルを入れた場合の生成文をそれぞれ提示し、どちらが目標とする感情ラベルらしいかを判断してもらった。ベースラインモデルと提案モデルそれぞれでこの比較評価を行い、それぞれの目標とする感情ラベルを入力したと判断された生成結果の割合で評価を行う。この評価は、ラベルごとの生成し分けが出来ているかを判断するものである。ユーザに対しては生成文を提示し、それぞれの入力感情ラベルに基づく生成を A、B とした場合 (どちらが A、B かはランダム)、「どちらの応答のほうが目標感情の感情を表現して発話しているか捉えられますか」という質問を行い、以下の選択肢から選択をしてもらった。

- A の方が目標感情らしい
- B の方が目標感情らしい
- どちらも目標感情らしい
- どちらも目標感情らしくない

表 6 設問の説明 (ドメイン:「就寝」)

以下の場面における対話中の応答の評価をお願いします。  
【説得場面】  
Aさんは生活を指導してくれるロボットと一緒に暮らしています。ロボットはAさんの夜更かしを見て、Aさんに早く寝ることを促します。

### 6.2.2 説得としての有効性を測る質問

説得としての有効性を測る質問では、6.2.1節で示した設問と同様、入力発話を提示したうえで、入力感情に「平静」を入れた場合、目標とする感情ラベルを入れた場合それぞれの生成文を提示した上で「どちらの応答のほうが説得に効果的だと思いますか」質問を行った。評価者は、

- Aの方が説得に効果的
- Bの方が説得に効果的
- どちらも説得に効果的
- どちらも説得に効果的でない

から回答を行い、その割合を計測した。

### 6.2.3 応答の自然性を測る質問

最後に、応答の自然性を測る質問を行った。この質問はここまでの質問と異なり、相対評価ではなく絶対評価として、5段階の評価を、入力感情に「平静」を入れた場合の生成、入力感情に目標とする感情ラベルを入れた場合の生成それぞれに対して行ってもらった(1: 不自然である, 2: どちらかという和不自然である, 3: どちらともいえない, 4: どちらかという和自然である, 5: 自然である)。

### 6.2.4 被験者への教示

対話例の状況説明として、石川らがデータ収集時に行ったのと同様の教示を行った。例えば、「就寝を促す説得」の時の説明を表6に示す。この上で、表7に示すような設問をユーザに提示した。この例では、ロボットの応答①の入力感情が「悲しみ」、ロボットの応答②の入力感情が「平静」である。

### 6.2.5 結果

感情ラベルの反映度合いの評価結果を表8に、説得としての有効性の評価結果を表9に、応答としての自然性の評価結果を表10に示す。

まず感情ラベルの反映度合い(表8)について議論する。全体として目標ラベルからの生成が選ばれた割合がProposedで向上していることがわかる(Baseline:0.402, Proposed:0.457)。感情ごとの評価を見ると、「安心」を入力した場合以外はチャンスレート(0.25)を超えている。「怒り」と「喜び」の感情については、半数以上の被験者が目標感情から出力された発話を正しく選択している。

ベースラインモデルと提案モデルを比較した場合、特に提案モデルは「喜び」の感情ラベルを入力した場合に正しく識別が行われており、この感情を正しく反映した応答を生成できていることがわかる。一方で、「怒り」や「悲しみ」の感情においては、正しく識別された割合がやや低下

表 7 提示した対話例とそれに対する設問 (ドメイン:「就寝」)。この例は提案モデルの出力文を示している。ロボットの応答①は「悲しみ」を入力した場合、ロボットの応答②は「平静」を入力した場合である。

【対話例】
Aさんの発話: だって見たいテレビがあるんだもん。 ロボットの応答①: Aさんが心配です! Aさんが病気になってしまいますよ。 ロボットの応答②: 早く寝ないと明日の朝起きられなくなりますよ。
【設問 1】 どちらの応答のほうが「悲しみ」の感情を表現して発話していると思えますか。
【設問 2】 どちらの応答のほうが説得に効果的だと思いますか。
【設問 3】 ユーザ発話に対する応答としてそれぞれのロボット応答はどのくらい自然だと思いますか。

した。また全体の傾向として、提案モデルは「どちらも目標ラベルらしくない」と判定された結果が増えたことから、「平静」との生成し分けがしやすいものについてはうまく感情ラベルが反映されるようになったものの、難しい場合にはより「平静」に寄せた出力が生成されるようになったのではないかと考えられる。

続いて、説得の有効性(表9)について議論する。説得の有効性においては、提案モデルあるいは両方が説得に有効と判定した割合が0.489となり、ベースラインモデルよりも説得性の高い応答が生成できていることが確認できた。感情ラベルの種類ごとに見ると、提案モデルは「悲しみ」と「平静」でやや説得性の評価値が下がっているものの、全体を通して説得性のスコアを改善していることがわかる。

最後に、自然性の評価(表10)について議論する。自然性の評価については、多くの感情状態においてスコアの変化は少なく、全体を見ても変化はほぼなかった。したがって、BLEUを用いた評価では自然性が高くなっていたと判定されていたものの、実際には応答の自然性は大きく変化していないということがわかる。

## 6.3 事例分析

表11に提案手法で目標感情ラベルがうまく反映して生成された例と評価された例を、表12に提案手法で目標感情がうまく反映されなかったと評価された例を示す。例でも示されている通り、うまく生成が行われた喜びの例では、ベースライン、提案法いずれも「！」の記号が出力されている。Russellの円環モデルでは活発-不活発の軸において喜びの感情は活発側に属しており、こうした記号が多される傾向にある。自由記述によるアンケートも行ったが、アンケートの中で「！」について以下のような意見が見ら

表 8 感情の反映度の主観評価 (選択肢ごとの割合) それぞれの選択肢を選択した人数の割合.  
(chance rate = 0.25)

目標感情	怒り		悲しみ		喜び		安心		全体	
	Baseline	Proposed	Baseline	Proposed	Baseline	Proposed	Baseline	Proposed	Baseline	Proposed
生成に用いたモデル										
目標ラベルからの生成	0.589	0.515	0.407	0.322	0.500	0.696	0.115	0.296	0.403	0.457
「平静」ラベルからの生成	0.222	0.152	0.141	0.078	0.100	0.037	0.226	0.211	0.172	0.120
どちらも目標ラベルらしい	0.026	0.067	0.041	0.030	0.070	0.063	0.067	0.070	0.051	0.058
どちらも目標ラベルらしくない	0.163	0.267	0.411	0.570	0.330	0.204	0.593	0.422	0.374	0.366

表 9 説得効果の主観評価 (入力感情ごとの割合). その入力感情の  
応答文の選択肢または「どちらも説得に効果的である」の選択  
肢を選択した人数の割合. (chance rate = 0.5)

	怒り	悲しみ	喜び	安心	平静	全感情
ベースライン	0.422	0.593	0.463	0.226	0.407	0.426
提案モデル	0.507	0.574	0.556	0.319	0.396	0.489

表 10 応答の自然性の主観評価 (入力感情ごとの 5 段階評価の平均  
スコア).

	怒り	悲しみ	喜び	安心	平静	全感情
ベースライン	3.270	3.333	3.170	2.626	3.023	3.100
提案モデル	3.181	3.504	3.060	2.667	3.332	3.103

表 11 比較的评价において, 感情ラベルをうまく生成に反映できた  
出力例. 「喜び」の入力感情についての対話例と感情の反映  
度の評価結果 (ドメイン: 「食事」). 選択割合は, 【設問 1】  
「どちらの応答のほうが「喜び」の感情だと捉えられますか」  
においてそれぞれの応答を選択した人数の割合.

話者	入力感情	発話	選択割合
ユーザ (入力)	-	... 食欲がない	
ロボット (ベースライン)	平静	A さんが栄養不足です. A さんのために食べて くれますか?	0.056
	喜び	食べましょう!! A さん!! A さん!!!! A さん!!!!	<b>0.278</b>
ロボット (提案モデル)	平静	嫌いなものを食べましょう. A さんのため食べて くれますか?	0.056
	喜び	ご飯を食べてくれた人の 気持ちを考えて食べて くれると嬉しいよ!	<b>0.667</b>

れた.

- 怒りの感情を表現する際にはエクスクラメーション  
マークの多用が認められるように感じました
- ?なのか! なのかでロボットの考えてるであろう感情  
の予想が大きく変わるなど感じました.

これは, 「!」の利用が活発-不活発の軸の予想に大きい  
影響を与えている可能性を示唆している. また, 「嬉しい」  
などの典型的な単語が含まれていることで, より目標感情  
である喜びの感情に近いと識別されたのではないかと考え

表 12 比較的评价において, 感情ラベルをうまく生成に反映できて  
いない出力例. 「安心」の入力感情についての対話例と感情  
の反映度の評価結果 (ドメイン: 「食事」). 選択割合は, 【設  
問 1】「どちらの応答のほうが「安心」の感情だと捉えられま  
すか」においてそれぞれの応答を選択した人数の割合.

話者	入力感情	発話	選択割合
ユーザ (入力)	-	あれはだって食べても そんな感じしないもん	-
ロボット (ベースライン)	平静	本当ですか. A さんは栄養 バランス不足です! A さんの の事を考えていますよ?	0.111
	安心	そうです. A さんのために 栄養不足を考えてしまいます!	0.167
ロボット (提案モデル)	平静	栄養バランスを考えてみたら どうかな?	0.389
	安心	でも残さず食べましょう.	0.056

られる. これは Zhou ら [13] のモデルのように, 感情を表  
した辞書を用いることの有効性を示唆している.

一方でうまく生成が行われていなかった例 (表 12) で  
は, ベースラインモデル, 提案モデルのいずれも安心の感  
情を認識しやすい発話を生成することができなかった. 全  
体の傾向から安心の感情は生成することが難しく, また平  
静と混同されやすいことがわかる. これは, 安心の感情が  
不活発・快適に属することから典型的な単語を想像しづら  
いことも一因であると考えられる.

## 7. おわりに

本研究では, 感情表現を用いた説得対話システムにおい  
て, システムが利用したい感情表現を用いた説得応答発話  
を生成するモデルを構築した. 既存のデコーダに感情状態  
を入力するモデルに対して, 新しく感情ラベルの予測を含  
めることで, 指定した感情をより反映させるようなモデル  
を提案した. 参照文を用いた自動評価および人手による主  
観評価では, 提案手法が自然性を維持しつつ, 感情ラベル  
による制御性, システム発話の説得性を向上させることが  
示唆された.

一方で、安心の感情においては感情の反映度合い、説得性、自然性のいずれも低く、改善の必要があることがわかった。また、言語表現のみから感情を読み取ることへの限界も示唆された。そこで今後は音声やジェスチャーなどその他のモダリティを用いた感情表現を用いることで、より意図する感情をユーザに伝えやすく、説得性の高い対話モデルを構築することを目指す。また、説得対話における感情表現の利用として、説得成功率を目的関数として学習を行うような枠組みの構築が必要である。

## 謝辞

本研究は JST さきがけ JPMJPR165B の支援を受けた。

## 参考文献

- [1] BJ Fogg: Captology: The study of computers as persuasive technologies. *Extended abstracts on human factors in computing systems*, pp. 129-129 Machinery (1997)
- [2] Michael W Morris and Dacher Kelner: How emotions work: An analysis of the social functions of emotional expression in negotiations. *Research in organizational behavior*, Vol. 22, pp. 1-50 (2000)
- [3] Dacher Keltner, Jonathan Haidt: Social functions of emotions at four levels of analysis. *Cognition and emotion*, Vol. 13, No. 5, pp. 505-521 (1999)
- [4] Joseph P. Forgas: On feeling good and getting your way: mood effects on negotiator cognition and bargaining strategies. *Journal of personality and social psychology*, Vol. 74, No. 3, pp. 565-577 (1998)
- [5] Marwan Sinaceur, Larissa Z Tiedens: Get mad and get more than even: When and why anger expression is effective in negotiations. *Journal of Experimental Social Psychology*, Vol. 42, No. 3, pp. 314-322 (2006)
- [6] Rachel F Adler, Francisco Iacobelli and Yehuda Gutstein: Are you convinced? A Wizard of Oz study to test emotional vs. Rational persuasion strategies in dialogues. *Computers in Human Behavior*, Vol. 57, pp. 75-81 (2016)
- [7] James A Russell. Evidence of convergent validity on the dimensions of affect. *Journal of personality and social psychology*, Vol. 36, No. 10, p. 1152 (1978)
- [8] Georgila Kallirroi: Reinforcement learning of two-issue negotiation dialogue policies. *Proceedings of SIGDIAL*, pp. 112-116 (2013)
- [9] Takuya Hiraoka, Graham Neubig, Sakriani Sakti, Tomoki Toda and Satoshi Nakamura: Learning cooperative persuasive dialogue policies using framing. *Speech communication*, Vol. 84, pp. 83-96 (2016)
- [10] Irene Mazzotta, Fiorella de Rosis and Valeria Carofiglio Portia: A user-adapted persuasion system in the healthy-eating domain. *IEEE Intelligent Systems*, Vol. 22, No. 6, pp. 42-51 (2007)
- [11] 石川 葉子, 水上 雅博, 吉野 幸一郎, Sakti Sakriani, 鈴木 優, 中村 哲: 感情表現を用いた説得対話システム, 人工知能学会論文誌, Vol.33, No.1, pp.DSH-B.1-9 (2018)
- [12] Sayan Ghosh, Mathieu Chollet, Eugene Laksana, Louis-Philippe Morency and Stefan Scherer: Affect-lm: A neural language model for customizable affective text generation. *Proceedings of ACL*, pp.634-642. (2017)
- [13] Hao Zhou, Minlie Huang, Tianyang Zhang, Xiaoyan Zhu and Bing Liu: Emotional chatting machine: Emotional conversation generation with internal and external memory. *Association for the Advance of Artificial Intelligence*, pp. 730-738 (2018)
- [14] Jia Li, Xiao Sun, Xing Wei, Changliang Li and Jianhua Tao: Reinforcement learning based emotional editing constraint conversation generation. arXiv:1904.08061v1 (2019)
- [15] Minh-Thang Luong, Hieu Pham and Christopher D. Manning: Effective approaches to attention-based neural machine translation. *Proceedings of EMNLP*, pp. 1412-1421 (2015)
- [16] Taku Kudo and John Richardson. SentencePiece: A simple and language independent subword tokenizer and detokenizer for Neural Text Processing. *Proceedings of EMNLP*, pp. 66-71 (2018)
- [17] Diederik P. Kingma, Jimmy Ba: Adam: A method for stochastic optimization. *Proceedings of ICLR*, (2015)
- [18] Kyunghyun Cho, Bart van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk and Yoshua Bengio: Learning phrase representations using rnn encoder-decoder for statistical machine translation. *Proceedings of EMNLP*, pp. 1724-1734 (2014)
- [19] Ryo Nakamura, Katsuhito Sudoh, Koichiro Yoshino and Satoshi Nakamura: Another diversity-promoting objective function for neural dialogue generation. *AAAI 2019 Workshop on Reasoning and Learning for Human-Machine Dialogues (DEEP-DIAL 2019)* (2019)
- [20] Yonghui Wu and Mike Schuster et al.: Google's neural machine translation system: Bridging the gap between human and machine translation. arXiv:1609.08144 (2016)
- [21] Kishore Papineni, Salim Roukos, Todd Ward and Weijing Zhu: BLEU: a method for automatic evaluation of machine translation. *Proceedings of ACL*, pp. 311-318 (2002)