Vol.2019-NL-243 No.15
2019/12/5

# 個別の発話スタイルを強調する Boosting Framework を用いた感情表現生成

尾関 晃英[1,a)]　李 晃伸[1]

**概要**：ニューラル対話システムにおいてより自然な対話を実現するために，感情や個性といった表現スタイルを扱える文生成手法が研究されている．特定のスタイルに沿った文生成を行うには，そのスタイルや話者等，目的に合ったラベル付きの発話データが必要であるが，目的やスタイルごとにデータを大量に取得，作成することは困難である．本研究では，小規模のラベル付き対話データとラベル無し大規模対話データからその特徴を強調した文生成モデルを効率的に学習するため，発話識別器を組み込んだ応答文生成手法を提案する．この発話識別器を本研究では Boosting Framework と呼ぶ．Boosting Framework は小規模のラベル付きデータから学習することで構築され，文生成モデルの学習時に組み込まれることで損失関数に対して目標スタイルに沿った重み付けを行う．感情ラベル付きデータを用いて感情表現を含む応答文生成をタスクとした実験において，比較手法では語彙の多様性と発話スタイルの両立が困難であった一方で，提案手法は多様性を大きくを損なうことなく発話スタイルを表出できることが示された．

# Emotional Response Generation Using Boosting Framework to Emphasize Specific Utterance Style

## 1. はじめに

近年，ソーシャルネットワーキングサービスなどから得た大規模な対話データを用いた，ニューラルネットワークに基づく対話システムを構築する研究が盛んである．中でも sequence-to-sequence（seq2seq）[1] は，入力された文脈に沿った妥当な応答を生成することができ，文生成モデルとして使用されることが多い．しかし，“I don’t know.” のようなあらゆる発話に通ずる応答，いわゆるダルレスポンスの頻出や，文脈との関連が低い応答，あるいは文法が破綻した不自然な応答の生成が問題視されている．これらの課題に対して，Wikipedia のような事実情報データを外部知識として用いる手法 [2][3] や応答のスタイルに着目したアプローチがあり，特に後者では，話者を表すベクトルや ID といった話者情報やラベル付きデータを用いる手法 [4][5][6][7]，特定のスタイルを持つ発話を収集したコーパスを用いて転移学習を行う手法 [8] などが提案されている．

任意の発話スタイルの再現には，そのスタイルを持つ発話や人手によるラベルが付与されたデータセットを用いることが望ましい．対話システムでは様々な発話スタイルの再現が求められるが，その多様な発話スタイルごとに大量のラベル付きデータセットを構築することは現実的には困難である．

そこで，小規模のラベル付き対話データとラベル無し大規模対話データからその特徴を強調した文生成モデルを効率的に学習するため，発話識別器を組み込んだ応答文生成手法を提案する．この発話識別器を本研究では Boosting Framework と呼ぶ．Boosting Framework は小規模のラベル付きデータから学習することで構築され，文生成モデルの学習時に組み込まれることで損失関数に対して目標スタイルに沿った重み付けを行う．これにより，学習データ中に潜在的に存在するスタイル表出度（スタイル性）の高い発話の効果的な学習が期待される．実験では目標スタイルを感情表現として応答文生成を行い，seq2seq や Boosting Framework に基づく学習データの事前フィルタリングを行う手法などと比較する．

本稿では生成文の自然性，多様性に関する自動評価および，文脈に対する関連性やスタイル性に関する人手評価の結果報告とその考察，分析について述べる．

[1] 名古屋工業大学
Nagoya Institute of Technology
[a)] ozeki@slp.nitech.ac.jp

## 2. 関連研究

ニューラルネットワークに基づく文生成モデルとして代表的な例に seq2seq[1] が挙げられる。seq2seq は任意の文脈を入力として扱うことが可能である一方で，ダルレスポンスを頻出する傾向は大きな課題とされている．対話はひとつの文脈に対してあらゆる応答が許容される one-to-many な性質と，逆に複数の文脈に対して共通の応答が許容される many-to-one な性質がある．したがってデータ全体を平均的に学習した結果，このようなモデルは過度に一般化された応答を生成しやすくなるとされている [9].

これに対するアプローチとしてはまず，モデルに対する入力に付加情報を用いる手法が挙げられる．具体的には，Wikipedia のような事実情報データを外部知識として用いる手法 [2][3] や文脈として対話履歴全体を入力とする手法 [10][11]，キャラクター性を表す話者ベクトルや話者 ID，感情ラベル付きデータを用いた手法 [4][5][6][7] などである．より多くの情報を基にして応答生成を行うことで生成文に情報性や発話スタイルを与え，過度な一般化を回避する．これらの手法に共通する懸念要素は，必要とする付加情報データは常に得られるわけではなく，また得られたとしても小規模である場合が多いということである．

また，学習データの質に問題があると考え，ノイジーデータに着目したアプローチもある．データフィルタリングによって学習データを事前に調整する手法では，Xu ら [12] は文脈と応答の類似性に基づいて，Purgai ら [9] は文脈と応答双方に対して許容され得る発話数の多さをその対話の entropy として算出し，これに基づいてフィルタリングを行った．このような手法は，人間の解釈に沿ってデータセットを調整することができる反面，モデルにとってのノイズの適切な定式化が難しく，その最適なフィルタリング強度も事後的に判断することとなる．一方で，ドメイン適応タスクにおいて用いられた instance weighting[13] を文生成モデルの学習に組み込んだ手法も提案されている [14]．この研究では，機械学習に基づく発話識別器 Calibrator による識別結果を基に，文生成モデルの損失関数に重み付けを行っている．ノイズの識別と学習への反映を一連の枠組みとしたこの手法では，ダルレスポンスを回避し，通常の seq2seq モデルよりも関連性の高い応答が生成できたことが報告されている．

以上より，本研究は Shang らの研究 [14] を発話スタイルの付与を目的として拡張することで，小規模データの活用と学習データの質に着目した文生成を試みる．

## 3. 個別の発話スタイルを強調する文生成モデル

提案モデルの概観を図 1 に示す．提案モデルは大きく分

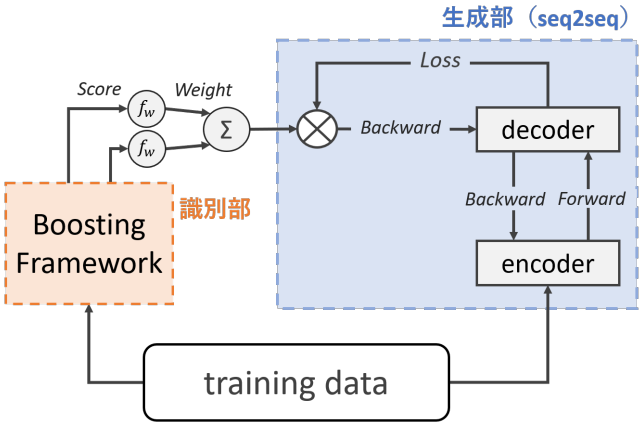


図 1　提案モデル概観.

けて生成部と識別部の 2 つで構成される．ここで，生成部では文生成モデルとして使用される attention 機構付き seq2seq を，識別部では発話スタイルの強調を行う発話識別器 Boosting Framework を使用する．提案モデルでは，seq2seq 学習時に Boosting Framework が各学習データに対して関連性，スタイル性に関する 2 つのスコアを算出し，これら基づいて seq2seq の損失に対して重み付けを行う．

以下，学習データを文脈 $X = \{x_1, x_2, ..., x_{T_X}\}$，正解応答 $Y = \{y_1, y_2, ....y_{T_Y}\}$（$T_X$，$T_Y$ は単語数）と表し，seq2seq の出力を $O = \{o_1, o_2, ..., o_{T_O}\}$（$T_O$ は単語数）と表す．

### 3.1 attention 機構付き seq2seq モデル

本研究における seq2seq モデルは encoder が 2 層の双方向 Gated Recurrent Unit（GRU），decoder が 2 層の順方向 GRU で構成される．encoder への入力として文脈 $X$ が与えられたとき，時刻 $t$ における入力単語 $x_t$ の単語分散表現を $e_t$ とすると，encoder の隠れ状態は

$$h_t = [\mathrm{GRU}(\overrightarrow{h_{t-1}}, e_t); \mathrm{GRU}(\overleftarrow{h_{t-1}}, e_t)] \quad (1)$$

と表され，最終時刻 $t = T_X$ における隠れ状態が文脈 $X$ を表す発話ベクトル $h_{con}$ となる．ここで，$\overrightarrow{h_t}$，$\overleftarrow{h_t}$ は時刻 $t$ における順方向，逆方向の隠れ状態を表し，$[;]$ はベクトルの連結を表す．

decoder では，encoder から得られた発話ベクトル $h_{con}$ を初期状態 $s_0$ とし，特殊記号＜ SOS ＞を初期入力単語 $o_0$ として順方向 GRU によって時刻 $u$ における隠れ状態 $s_u$ を以下のように算出する．

$$s_u = \mathrm{GRU}(s_{u-1}, o_{u-1}) \quad (2)$$

decoder による単語予測を行う際に文脈との対応関係をより考慮したモデリングを行うため，attention 機構を導入する．attention 機構は decoder において $u$ ステップ目の処理を行うとき，decoder の隠れ状態 $s_u$ と encoder の各隠れ状態 $h_t (t = 1, 2, ..., T_X)$ のスコア $a_u(h_t)$ を計算する．その後，これらのスコアに基づいて encoder の隠れ状態を

統合したコンテキストベクトル $c_u$ を算出し，$s_u$ と組み合わせることで新たな隠れ状態 $s'_u$ とする．

$$a_u(h_t) = v_a^{\mathsf{T}} \tanh\left(W_a[s_u; h_t]\right) \tag{3}$$

$$c_u = \sum_{t=1}^{T_X} \mathrm{softmax}([a_1, a_2, ..., a_{T_X}])h_t \tag{4}$$

$$s_u^{'} = \tanh(W_c[c_u; s_u]) \tag{5}$$

最後に，attention 機構で得られた $s'_u$ を用いて次式のように出力単語 $o_u$ を予測する．

$$o_u = \mathrm{softmax}(Ws'_u) \tag{6}$$

なお，$v_a$，$W_a$，$W_c$，$W$ は学習パラメータである．

### 3.2 Boosting Framework

Boosting Framework の構造を図 2 に示す．Boosting Framework は文脈および正解応答に対する 2 つの encoder と多層パーセプトロン（MLP）によって構成される発話識別器であり，入力された発話ペア $(X, Y)$ の関連性および $Y$ のスタイル性に関するスコアを出力する．処理としては，まず単層の双方向 GRU によって seq2seq の encoder と同様の手順で文脈 $X$，正解応答 $Y$ をそれぞれの発話ベクトル $h_{con}$，$h_{res}$ にエンコードする．次に，$h_{con}$，$h_{res}$ を用いて $(X, Y)$ の関連性に関するスコア $s_{rel}$ および $Y$ のスタイル性に関するスコア $s_{sty}$ を得る．ここで，$s_{sty}$ は $Y$ が目標とする発話スタイルの特徴を強く持っていると識別されるほど大きな値となる．

$s_{rel}$ は Tao らの提案した評価尺度 [15] で用いられた Unreferenced score を使用し，次式で求められる．

$$s_{rel} = \mathrm{MLP}([h_{con}; h_{con}^{\mathsf{T}} W_r h_{res}; h_{res}]) \tag{7}$$

$W_r$ は学習パラメータを，$MLP()$ は 2 層の MLP への入力を表しており，MLP では活性化関数として 1 層目で tanh 関数，2 層目で sigmoid 関数を用いる．

また，$s_{sty}$ は次式で求められる．

$$s_{sty} = \mathrm{MLP}(h_{res}) \tag{8}$$

### 3.3 Boosting framework を用いた文生成

3.1 節で示した attention 機構付き seq2seq モデルの学習では，次式で表される負の対数尤度 $L_{nll}$ の最小化を行う．

$$L_{nll} = -\frac{1}{T_O} \sum_{t=1}^{T_O} logP(o_t) \tag{9}$$

ここで，提案モデルでは図 1 のように，Boosting Framework から出力された $s_{rel}$，$s_{sty}$ を変換式 $f_w$ によって重み $w_{rel}$，$w_{sty}$ に変換したのち，その加重平均 $w$ を用いて損失関数に対して重み付けを行う．したがって，最終的な損失は

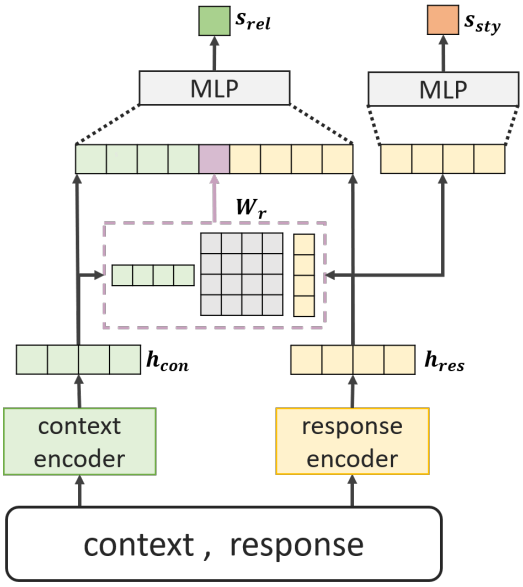


図 2　Boosting Framework の構造．

$$L_{weighted\_nll} = wL_{nll} \tag{10}$$

と表される．

バッチサイズ $b$ のミニバッチにおける $i$ 番目のデータの $s_{rel}$，$s_{sty}$ をそれぞれ $s^i_{rel}$，$s^i_{sty}$ とし，$s^i_{rel}$，$s^i_{sty}$ を変換した重みを $w^i_{rel}$，$w^i_{sty}$ とすると，変換式 $f_w$ は

$$w^i_{rel(sty)} = f_w(s^i_{rel(sty)}) = \frac{s^i_{rel(sty)}}{\frac{1}{b}\sum_{i=1}^{b} s^i_{rel(sty)}} \tag{11}$$

と表され，$w^i_{rel}$，$w^i_{sty}$ の加重平均 $w^i$ は重み係数 $\gamma$ によって

$$w^i = (1-\gamma)w^i_{rel} + \gamma w^i_{sty} \tag{12}$$

と表される．$s^i_{rel(sty)}$ が [0,1] で値をとる一方，変換された重み $w^i_{rel(sty)}$ はミニバッチ内の $s_{rel(sty)}$ の平均スコアに対する $s^i_{rel(sty)}$ の比であるため，各ミニバッチ処理ごとに $w^i_{rel(sty)} > 1$ もしくは $w^i_{rel(sty)} < 1$ となる重みがそれぞれ $\frac{2}{b}$ 個ずつ出力されることとなる．以上より，提案モデルでは関連性およびスタイル性がミニバッチ内で相対的に高いデータの損失に基づいてパラメータ更新が行われる．

## 4. Boosting Framework の構築と評価

Boosting Framework を構築するためにラベル付き小規模データを用いて学習を行う．

### 4.1 学習

ある（文脈，応答）ペアに対して，関連性とスタイル性に関する 2 値ラベル $l_{rel}$，$l_{sty}$ が付与されているとき，Boosting Framework はその文脈と応答から得た $s^i_{rel}$，$s^i_{sty}$ から次式の損失 $L_{BF}$ に基づいて 2 種類のラベル予測タスクを同時に学習する．ここで，$\alpha$ は各タスクの重みを調整

表 1　DailyDialog データセットの詳細.

| | train | dev | test |
|---|---|---|---|
| 対話数 | 13111 | 1000 | 1000 |
| 発話ペア数 | 54780 | 5144 | 3532 |

表 2　train，dev データにおけるラベルパターンの内訳.

| パターン | 関連 | 感情 | train | dev | test |
|---|---|---|---|---|---|
| a | ○ | ○ | 13695 | 1286 | 883 |
| b | × | ○ | 13695 | 1286 | 883 |
| c | ○ | × | 13695 | 1286 | 883 |
| d | × | × | 13695 | 1286 | 883 |

する係数である．

$$L_{BF} = (1-\alpha)\text{Binary\_cross\_entropy}(l_{rel}, s_{rel}) + \alpha\text{Binary\_cross\_entropy}(l_{sty}, s_{sty}) \quad (13)$$

### 4.2　データセット

Boosting Framework 学習用のデータセットとして DailyDialog[16] を使用した．これは英語学習者向け web サイトから日常会話例を収集したものであり，各対話は平均 7.9 ターンで構成されている．今回はこれらの対話を 2 発話ごとに (文脈，応答) ペアとして分解し再構築した．使用したデータセットの詳細を**表 1** に示す．

各発話には感情ラベル（Anger，Disgust，Fear，Happiness，Sadness，Surprise，Other）と発話行為ラベルが人手で付与されており，各対話に対しては話題ラベルが付与されているが，対話行為ラベル及び話題ラベルは本研究では使用しない．さらに，Anger，Disgust，Fear，Happiness，Sadness，Surprise の 6 種類の感情ラベルを “感情あり” として 1 つのラベルに統合し，Other を “感情なし” とすることで感情表現の有無を表す二値のラベルとして使用する．これは本研究では感情表現という発話スタイルの強調の際，その感情の種類は問わないためである．内容の関連性に関するラベルは，DailyDialog の実際の発話ペアを “関連あり” とし，元の対話が異なる別の発話ペアからランダムにサンプリングした発話を応答としたペアを “関連なし” として新たに作成した．これは，人手で整形された DailyDialog は文脈と応答の関連性が全体的に高い一方で，大規模に収集した対話データには関連性が低いノイジーデータが含まれることが想定されるためである．以上より，発話ペアのラベルパターンは**表 2** に示す 4 パターンが存在する．ここで，パターン b，d はそれぞれパターン a，c と同じ文脈である．発話ペア作成の際，パターン a の数に合わせて他のラベルパターンを作成した．

### 4.3　モデル設定

Encoder の隠れ層は次元数を 128 とし，単語ベクトルには 100 次元の GloVe[17] を用いた．モデルが使用する語彙は 5.3 節で述べた文生成モデルと同じものを用いる．学習パラメータの最適化には Adam[18] を使用し，学習率を 0.0001 とした．また，識別性能が最良のモデルを判断するため，式（13）における重み係数 $\alpha$ を 0.0 から 1.0 まで 0.1 刻みで変化させてそれぞれ学習を行った．これらのモデルは最大 20 epoch の学習を行い，1 epoch 毎に算出した開発ロスが最も低いモデルで評価を行った．

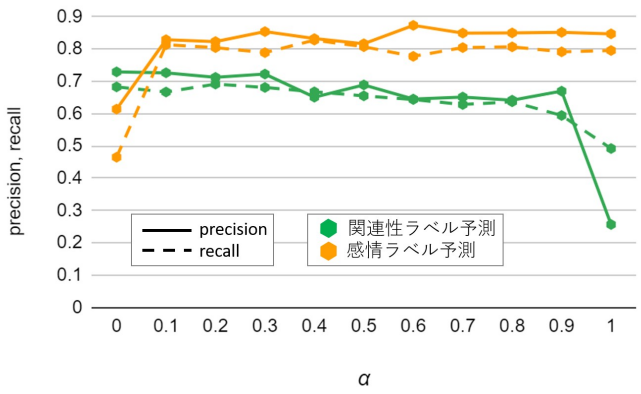


図 3　各 $\alpha$ における precision および recall．実線が precision，破線が recall を表す.

### 4.4　評価とモデル選定

各 $\alpha$ におけるテストデータに対する precision，recall を図 3 に示す．図 3 より，$\alpha$ の値を大きくするにつれて関連性ラベル予測に対する精度が下がる結果となった．これは $\alpha$ を大きくすることで感情ラベル予測に大きな比重を置いた学習を行うようになるためであり，逆に $\alpha = 0.0$ の場合に感情ラベル予測の精度が大きく下がっている．

関連性ラベル予測と感情ラベル予測の $F$ 値に基づいて最良のモデルを決定した結果，以降で使用する Boosting Framework は $\alpha = 0.3$ のモデルとした．また，このモデルのテストデータに対する accuracy は関連性ラベル予測が 0.692，感情ラベル予測が 0.812 であった．なお，用いたデータセットにラベルの偏りがないことから accuracy 算出のための閾値は 0.5 としている．

## 5.　実験

文脈に対する応答文生成をタスクとした実験を行い，提案手法の有効性を検証する．なお本実験では目標スタイルを感情表現とする．

### 5.1　データセット

文生成モデルの学習データとして DSTC7-Task2[19] において使用された reddit データセットを用いる．この対話データは reddit で行われた 178 種の subreddit（特定トピックの議論を行う場）から収集され，前処理として Markdown と特殊記号の削除，および NLTK（Natural Language

表 3　reddit データセットの詳細.

| | train | dev | test |
|---|---|---|---|
| 対話数 | 1865654 | 40932 | 13440 |
| 発話ペア数 | 2256276 | 72222 | — |

Toolkit)[*1]の TweetTokenizer を用いた言語表現の統一を行った．また，複数ターンに渡る対話を 2 発話ごとに（文脈，正解応答）ペアとしてデータセットを再構築し，発話最大長は 60 単語とした．また，評価時にはテストデータの対話履歴における最終発話のみを入力して応答を生成した．再構築後の reddit データセットの詳細を**表 3** に示す．

### 5.2　比較モデル

本実験では Boosting Framework の使用方法として，あらかじめ関連性とスタイル性が高いデータを抽出するための事前フィルタリングに Boosting Framework を用いる手法と，文生成モデルの学習に直接組み込んで用いる提案手法を比較する．比較モデルを以下に示す．

- `S2SA`：3.1 節における attention 機構付き seq2seq モデル．本実験におけるベースラインとする．
- `S2SA+T`：reddit データセットで学習済みの `S2SA` を DailyDialog データセットを用いて転移学習したモデル．
- `S2SA+F(`$n$`%)`：`S2SA` において，Boosting Framework に基づく事前フィルタリングを施した学習データを用いたモデル．
- `S2SA+F+T(`$n$`%)`：`S2SA+F(`$n$`%)` を `S2SA+T` と同様の方法で転移学習したモデル．
- `S2SA+B`：3.3 節における Boosting Framework を `S2SA` の学習に組み込んだモデル．
- `S2SA+B+T`：`S2SA+B` を `S2SA+T` と同様の方法で転移学習したモデル．

Boosting Framework による学習データの事前フィルタリングでは，各発話ペアに対する $s_{rel}$ と $s_{sty}$ を算出し，その加重平均に基づいて，上位 $n$%のみを使用する（$n = 90, 80, ..., 10$）．この際，関連性とスタイル性に関して `S2SA+B` と同様の基準でフィルタリングを行うために，加重平均の重み係数は式（12）の $\gamma$ と同値に設定する．

転移学習の設定に関する詳細は次節で述べる．

### 5.3　モデル設定

encoder，decoder は隠れ層の次元数を 128 とし，encoder の入力となる単語ベクトルには 100 次元の GloVe を用いた．モデルが使用する語彙は reddit データセットの訓練データにおける出現回数上位 20000 単語から作成し，それ以外の単語は未知語を表す特殊記号＜ UNK ＞に変換した．学習パラメータの最適化には Adam[18] を使用し．学習率を 0.0005 とした．学習は最大 10 epoch 行い，1 epoch 毎に開発データから算出した perplexity が最も低いモデルで評価する．また，文生成には top-$k$ random sampling[20] を使用した．以上を比較モデル共通の設定とする．

`S2SA+T`，`S2SA+F+T(`$n$`%)`，`S2SA+B+T` の学習パラメータの初期値は，それぞれ転移学習前のモデルにおいて評価に用いることとされたモデルの値を使用する．転移学習には表 2 に示したパターン a の発話ペアを用いて最大 10epoch 学習した．

また，`S2SA+B` のハイパパラメータ $\gamma$ は 0.5 とした．

### 5.4　評価方法

テストデータに対する生成文を用いて自動評価と人手評価を行う．自動評価尺度は以下に示す 5 つである．

**BLEU**　生成文と正解応答との類似度を $N$-gram（$N = 3, 4$）一致率を基に算出する（BLEU-3,4）．生成文の語順の正しさに着目した指標である．

**NIST**　生成文と正解応答との類似度を $N$-gram（$N = 3, 4$）一致率を基に算出する（NIST-3,4）．個々の $N$-gram に対して情報量に基づく重み付けがされるため，生成文の文意に着目した指標であるといえる．

**div**　生成文におけるユニークな $N$-gram（$N = 1, 2$）の数を全ての $N$-gram の数で除算することで算出する（div-1,2）．モデルが生成した $N$-gram の多様性を表す指標である．

**Valance**　BLEU-4, NIST-4, div-1 の相乗平均．すなわち，

$$\text{Valance} = \sqrt[3]{\text{BLEU-4} \times \text{NIST-4} \times \text{div-1}} \qquad (14)$$

Zhao ら [21] らは複数の評価尺度の相乗平均をとることで，その全体的な中庸さを表す指標とした．本研究における Valance はこれを参考にした尺度である．

自然言語処理タスクにおいて一般的に使用される BLEU や NIST，div は定められた正解文との類似度や多様性を計る尺度であり，発話スタイルや関連性の評価を直接的に行うものではない．また，非タスク指向対話では BLEU や NIST を含む word-overlap 指標は人手評価との相関が低いことが報告されている [22]．したがって，応答の直接的な評価を人手によって行う．

人手評価は以下に示す 3 つの項目に関し（0，1，2）の 3 段階評価を，計 4 名の評価者によって各モデル 50 サンプルに対して行う．評価項目と評価値およびその基準，または項目の説明を以下に示す．

**感情表出度**

**0**　感情が表出されていない
**1**　感情が表出されている
**2**　感情が強く表出されている

**文脈との関連性**

**0**　関連がない

[*1] http://www.nltk.org/

**1** 多少は関連がある
**2** 強い関連がある

**流暢さ**
**0** 文法が破綻している
**1** 許容できる
**2** 流暢で自然である

**感情表出率** 感情表出度において 1 または 2 と回答されたサンプルの割合．

また，評価者間一致を表す指標として $\kappa$ 値を算出した．

### 5.5 結果と考察

#### 5.5.1 自動評価

自動評価結果を**表 4** に示す．なお，S2SA+F は $n = 50, 10$ の結果のみを掲載した．表 4 より，発話スタイルを付与するすべてのモデルの div がベースライン (S2SA) と比較して減少する結果となった．発話スタイルの変化とはモデルが生成する語彙の傾向の変化であるため，比較モデルで唯一，一貫した発話スタイルを持たない S2SA の div が高いことは自然である．

加えて，発話スタイル付与を行ったモデルから転移学習を行ったモデル（S2SA+F+T(50%)，S2SA+F+T(10%)，S2SA+B+T）に注目すると，これらすべてのモデルで転移学習前（S2SA+F(50%)，S2SA+F(10%)，S2SA+B）より div がさらに減少している．以上の結果より，発話のスタイル性と生成する語彙の多様性はトレードオフの関係であるといえる．

S2SA+F(50%)，S2SA+B は Valance において最も高く，S2SA とほぼ同等であった．したがって，これらのモデルはいずれもスタイル性の増加により多様性の損失はあるものの，全体の自然性を保つことができたと考えられる．

#### 5.5.2 人手評価

人手評価結果を**表 5** に示す．転移学習を行ったすべてのモデル（S2SA+T，S2SA+F+T(50%)，S2SA+F+T(10%)，S2SA+B+T）の感情表出度および感情表出率に注目すると，それぞれ転移学習前のモデル（S2SA，S2SA+F(50%)，S2SA+F(10%)，S2SA+B）より上がっていたことから，赤間らの研究 [8] と同様に発話スタイル表出のための手法として転移学習は有効であることが示された．しかし，S2SA+T は S2SA と比べ関連性が下がっており，転移学習のみを行った単純な手法では不十分であることが示された．

S2SA+B は自動評価においては S2SA+F(50%) と同等の性能であったが，人手評価ではすべての指標で上回る結果となった．したがって，自動評価から推測されたように S2SA+B は自然性や多様性を大きく損なうことなく発話スタイルを表出できた上に，その内容もより適切なものであることが示された．

さらに，S2SA+B+T は S2SA+B に次いで関連性で高い評価となり，感情表出度および感情表出率においては比較

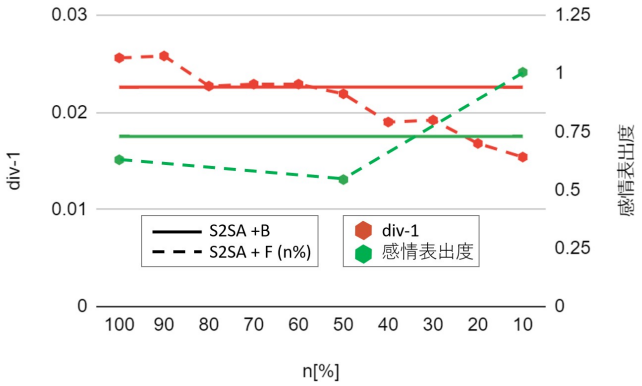


**図 4** S2SA+B と各 $n$ における S2SA+F の div および感情表出度の比較．実線は S2SA+B を表し，$n$ に依存しない．

モデル中で 1 番，流暢さにおいても全体で 2 番目に高い評価を得た．以上の結果より，転移学習を行う場合でもその事前学習に Boosting Framework を組み込んだ本手法は有効であることが示された．

### 5.6 分析

#### 5.6.1 提案手法と事前フィルタリング手法の比較

S2SA+F における $n$ を変化させたときの div-1 と感情表出度の推移を図 4 に示す．S2SA+F の div-1 を表す破線に注目すると，データをより多くフィルタリングするほど div-1 が減少する傾向が見られる．また，この減少傾向が強くなるのはデータ量がおよそ 110 万である $n = 50$ 以降の場合であった．

感情表出度は，$n = 50$ の場合と $n = 10$ の場合における差が大きいことが確認できる．div-1 の傾向と合わせると，フィルタリングによって学習データを厳選するほど発話スタイルの再現性は高まる一方，過度なフィルタリングは応答の偏りを招くこととなる．

S2SA+B は，各 $n$ における S2SA+F の中で比較的高い div-1 であった場合とほぼ同等の多様性でありながら，より高い感情表出度であったことが分かる．したがって，Boosting Framework は事前フィルタリングに用いる場合よりも，文生成モデルの学習に直接組み込む提案手法がより効果的である．

#### 5.6.2 各手法における語彙傾向の変化

次に，モデルがテストデータに対して生成した語彙傾向を調査した．図 5 にモデル間の頻出単語の差分を示す．なお，ここでは計 100 回以上生成された語彙を頻出単語としており，単語感情極性対応表 [23] において極性値が 0.1 以上または-0.1 以下の単語を色付き太字で示し，これらを感情語と定義する．S2SA+B - S2SA では，“sorry” や “thanks”，“awesome” などの感情語が実際に差分に含まれていたことから，発話スタイル付与に関して提案手法の有効性が確認された．

表 4　自動評価結果.

| Models | BLEU-3 | BLEU-4 | NSIT-3 | NSIT-4 | div-1 | div-2 | Valance |
|---|---|---|---|---|---|---|---|
| S2SA | 0.032 | 0.011 | 2.128 | 2.129 | **0.026** | **0.180** | 0.085 |
| S2SA+T | 0.024 | 0.007 | 1.201 | 1.201 | 0.017 | 0.137 | 0.052 |
| S2SA+F(50%) | 0.032 | **0.013** | 2.224 | 2.225 | 0.022 | 0.163 | **0.086** |
| S2SA+F(50%)+T | 0.024 | 0.006 | 1.442 | 1.442 | 0.017 | 0.150 | 0.053 |
| S2SA+F(10%) | **0.033** | **0.013** | 1.974 | 1.975 | 0.015 | 0.132 | 0.073 |
| S2SA+F(10%)+T | 0.021 | 0.005 | 1.233 | 1.230 | 0.013 | 0.122 | 0.043 |
| S2SA+B | **0.033** | 0.012 | **2.299** | **2.300** | 0.023 | 0.162 | **0.086** |
| S2SA+B+T | 0.023 | 0.006 | 1.076 | 1.076 | 0.018 | 0.153 | 0.049 |

表 5　人手評価結果.

| Models | 感情表出度 | 関連性 | 流暢さ | 感情表出率 [%] | $\kappa$ |
|---|---|---|---|---|---|
| S2SA | 0.630 | **0.620** | 1.080 | 44.5 | 0.157 |
| S2SA+T | 1.130 | 0.415 | **1.330** | 74.0 | 0.274 |
| S2SA+F(50%) | 0.545 | 0.485 | 0.840 | 43.0 | 0.146 |
| S2SA+F+T(50%) | 1.055 | 0.510 | 1.215 | 71.0 | 0.257 |
| S2SA+F(10%) | 1.005 | 0.485 | 0.950 | 68.5 | 0.302 |
| S2SA+F+T(10%) | 1.080 | 0.360 | 1.065 | 71.0 | 0.216 |
| S2SA+B | 0.730 | 0.550 | 1.070 | 54.0 | 0.144 |
| S2SA+B+T | **1.175** | 0.520 | 1.290 | **80.0** | 0.195 |

S2SA+T と S2SA の差分にはさらに多くの感情語が含まれていた．単語感情極性対応表に基づく感情語以外にも感嘆表現としての“oh”や”surprised”のような単語や，“looking forward”のように感情を表現するフレーズの単語が含まれていたことから，実際に多くの対話において感情表現を行っていたことがわかる．

S2SA+T と S2SA+B+T の比較では，頻出単語の差分は S2SA+T と S2SA の差分より少なく，感情表現のバリエーションや頻度に関して大きな差はなかった．実際，S2SA+T は“terrible”を 72 回生成しており，S2SA+B+T は“beautiful”を 91 回，“afraid”を 77 回生成していた．しかし自動評価および人手評価の結果を踏まえると，総合的により質の高い応答ができたのは S2SA+B+T であったといえる．

### 5.6.3　対話例

最後に，各モデルの実際の応答例を図 6 に示す．S2SA+B は“*best*”という強い極性を持つ単語を生成しており，S2SA や S2SA+F(50%) とは異なり感情表現により近い応答ができている．さらに，文脈における“*king cobra*”の話題に対し S2SA+F(50%)，S2SA+F(10%) では“*king*”や“*lions*”，“*dragons*”といった単語を生成しているが，S2SA+B では“*snakes*”のように直接的な関連がある単語を含んだ応答であった．

転移学習を行ったモデルは“*oh*”という感嘆詞や“*cool*”，“*love*”，“*interesting*”のような知覚を表す単語を生成しており，感情表現が表出されているといえる．また，S2SA+T や S2SA+F+T（50%）では文脈に対して不自然な応答内容となっている一方で，S2SA+B+T の内容は許容されるものである．

| Model1 | Model2 | Model1 – Model2 |
|---|---|---|
| S2SA+B | S2SA | too, **sorry**, oh, sounds, seen, case, **thanks**, might, probably, **awesome**, **cool**, rest, ... |
| S2SA+T | S2SA | seems, **ok**, **thanks** , surprised, **kind**, feel, **happy**, **beautiful**, kidding, **wonderful**, too, **thank**, **fun**, oh, forward, looking, **afraid**, **amazing**, surprise, sounds, wow, **nice**, appreciate, looks, **sorry**, ... |
| S2SA+B+T | S2SA+T | next, then, lets, again, still, happened, fine, now, **terrible**, ... |
| S2SA+T | S2SA+B+T | many, surprise, name, her, **beautiful**, heard, **afraid**, company, need, ... |

図 5　モデル間の頻出語彙の差分．色付き太字の単語は単語感情極性対応表 [23] において極性が 0.1 以上もしくは −0.1 以下である単語.

## 6.　おわりに

本研究では生成文への発話スタイル付与を目的としたラベル付き小規模データの効果的な応答文生成手法として，発話識別器 Boosting Framework を用いた文生成モデルを提案した．実験の結果，提案手法は多様性や関連性の大きな損失を抑えつつ発話スタイルの表出ができることが示され，転移学習を組み合わせる場合でも事前学習において提案手法を用いたモデルは人手評価において高い評価を得た．今後は Boosting Framework の識別に関する分析や，他の発話スタイルにおける同様の実験を行う予定である．また人手評価に関して，今回は非ネイティブによる評価を行ったが，感情表現の人手評価における非ネイティブ・ネイティブの差異についても今後は検討する必要がある．

### 参考文献

[1] Sutskever, I., Vinyals, O. and Le, Q. V.: Sequence to Sequence Learning with Neural Networks, *NIPS 2014, December 8-13 2014, Montreal, Quebec, Canada*, pp. 3104–3112 (2014).
[2] Ghazvininejad, M., Brockett, C., Chang, M., Dolan, B., Gao, J., Yih, W. and Galley, M.: A Knowledge-Grounded Neural Conversation Model, *AAAI, New Orleans, Louisiana, USA, February 2-7, 2018*, pp. 5110–5117 (2018).
[3] Dinan, E., Roller, S., Shuster, K., Fan, A., Auli, M. and Weston, J.: Wizard of Wikipedia: Knowledge-Powered Conversational agents, *CoRR*, Vol. abs/1811.01241 (2018).
[4] Li, J., Galley, M., Brockett, C., Spithourakis, G. P., Gao,

**Context :**
***the king cobra is unusual from other snakes in that it growls like an angry dog , instead of hisses .***

| Model | Response |
|---|---|
| S2SA | *well the king doesn't have a ring .* |
| S2SA+T | *oh , that's so cool ! i know a dog who likes mushrooms , even if they are unusual from animal control .* |
| S2SA+F(50%) | *this is what i've seen as the king .* |
| S2SA+F+T(50%) | *oh , honey , i love those . i am not sure that you're so shy .* |
| S2SA+F(10%) | *i have to watch some of the kings lions and dragons , and they were so scared . so i can find out why would they eat a cat .* |
| S2SA+F+T(10%) | *i like the jungle movie .* |
| S2SA+B | *the snakes are the best .* |
| S2SA+B+T | *oh , that sounds interesting .* |

**図 6** 各モデルの実際の応答例.

J. and Dolan, W. B.: A Persona-Based Neural Conversation Model, *ACL 2016, August 7-12, 2016, Berlin, Germany, Volume 1: Long Papers* (2016).

[5] Joshi, C. K., Mi, F. and Faltings, B.: Personalization in Goal-Oriented Dialog, *CoRR*, Vol. abs/1706.07503 (2017).

[6] Zhou, H., Huang, M., Zhang, T., Zhu, X. and Liu, B.: Emotional Chatting Machine: Emotional Conversation Generation with Internal and External Memory, *AAAI, New Orleans, Louisiana, USA, February 2-7, 2018*, pp. 730–739 (2018).

[7] Zhou, X. and Wang, W. Y.: MojiTalk: Generating Emotional Responses at Scale, *ACL 2018, Melbourne, Australia, July 15-20, 2018, Volume 1: Long Papers*, pp. 1128–1137 (2018).

[8] 赤間怜奈，稲田和明，小林颯介，佐藤祥多，乾健太郎，東北大学：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).

[9] Csaky, R., Purgai, P. and Recski, G.: Improving Neural Conversational Models with Entropy-Based Data Filtering, *ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers*, pp. 5650–5669 (2019).

[10] Serban, I. V., Klinger, T., Tesauro, G., Talamadupula, K., Zhou, B., Bengio, Y. and Courville, A. C.: Multiresolution Recurrent Neural Networks: An Application to Dialogue Response Generation, *AAAI, February 4-9, 2017, San Francisco, California, USA.*, pp. 3288–3294 (2017).

[11] Xing, C., Wu, Y., Wu, W., Huang, Y. and Zhou, M.: Hierarchical Recurrent Attention Network for Response Generation, *AAAI, New Orleans, Louisiana, USA, February 2-7, 2018*, pp. 5610–5617 (2018).

[12] Xu, X., Dusek, O., Konstas, I. and Rieser, V.: Better Conversations by Modeling, Filtering, and Optimizing for Coherence and Diversity, *EMNLP, Belgium, October 31 - November 4, 2018*, pp. 3981–3991 (2018).

[13] Wang, R., Utiyama, M., Liu, L., Chen, K. and Sumita, E.: Instance Weighting for Neural Machine Translation Domain Adaptation, *EMNLP 2017, Copenhagen, Denmark, September 9-11, 2017*, pp. 1482–1488 (2017).

[14] Shang, M., Fu, Z., Peng, N., Feng, Y., Zhao, D. and Yan, R.: Learning to Converse with Noisy Data: Generation with Calibration, *IJCAI 2018, July 13-19, 2018, Stockholm, Sweden.*, pp. 4338–4344 (2018).

[15] Tao, C., Mou, L., Zhao, D. and Yan, R.: RUBER: An Unsupervised Method for Automatic Evaluation of Open-Domain Dialog Systems, *CoRR*, Vol. abs/1701.03079 (2017).

[16] Li, Y., Su, H., Shen, X., Li, W., Cao, Z. and Niu, S.: DailyDialog: A Manually Labelled Multi-turn Dialogue Dataset, *IJCNLP 2017, Taipei, Taiwan, November 27 - December 1, 2017 - Volume 1: Long Papers*, pp. 986–995 (2017).

[17] Pennington, J., Socher, R. and Manning, C. D.: Glove: Global Vectors for Word Representation, *EMNLP 2014, October 25-29, 2014, Doha, Qatar, A meeting of SIGDAT, a Special Interest Group of the ACL*, pp. 1532–1543 (2014).

[18] Kingma, D. P. and Ba, J.: Adam: A Method for Stochastic Optimization, *ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings* (2015).

[19] Michel Galley, Chris Brockett, X. G. B. D. J. G.: End-to-End conversation Modeling: DSTC7 Task 2 Description, *DSTC7 workshop* (2019).

[20] Fan, A., Lewis, M. and Dauphin, Y. N.: Hierarchical Neural Story Generation, *ACL 2018, Melbourne, Australia, July 15-20, 2018, Volume 1: Long Papers*, pp. 889–898 (2018).

[21] Zhao, T. and Eskénazi, M.: Zero-Shot Dialog Generation with Cross-Domain Latent Actions, *SIGdial, Melbourne, Australia, July 12-14, 2018*, pp. 1–10 (2018).

[22] Lowe, R., Noseworthy, M., Serban, I. V., Angelard-Gontier, N., Bengio, Y. and Pineau, J.: Towards an Automatic Turing Test: Learning to Evaluate Dialogue Responses, *ACL 2017, Vancouver, Canada, July 30 - August 4, Volume 1: Long Papers*, pp. 1116–1126 (2017).

[23] Takamura, H., Inui, T. and Okumura, M.: Extracting Semantic Orientations of Words using Spin Model, *ACL 2005, 25-30 June 2005, University of Michigan, USA*, pp. 133–140 (2005).

# 正誤表

下記の箇所に誤りがございました．お詫びして訂正いたします．

| 訂正箇所 | 誤 | 正 |
|---|---|---|
| 1 ページ<br>概要 | 大きくを | 大きく |
| 5,6,7 ページ<br>各所 | Valance | Balance |
| 7 ページ<br>表 4 | S2SA+F(50%)+T,<br>S2SA+F(10%)+T | S2SA+F+T(50%),<br>S2SA+F+T(10%) |
| 7 ページ<br>表 4 | NSIT-3, NSIT-4 | NIST-3, NIST-4 |