Vol.2020-HCI-190 No.26
Vol.2020-UBI-68 No.26
2020/12/9

ER-Chat: 感情制御を伴う対話システムにおけるテキスト応答生成手法

片山 晋[1,a)]　米澤 拓郎[1]　大越 匡[2]　中澤 仁[3]　河口 信夫[1,4]

概要：対話システムにおいて感情表現を行うことは，ユーザの生活満足度の向上や肯定的なインタラクションの増加などに繋がる．近年ではニューラルネットワークを用いて感情的な応答生成手法も提案されているため，より自然で円滑な対話応答生成が可能である．しかし，既存研究における感情的な応答生成手法は，人間同士の対話コミュニケーションのような対話相手の感情状態や発話内容を考慮した上で応答時の感情を決定する感情制御を行っていない．そこで本研究では，ユーザの発話入力テキストから応答時の感情カテゴリを推定する感情制御を行い，感情的な対話応答を生成する ER-Chat を提案する．本手法は入力テキストから推定された応答時の感情カテゴリに即した応答テキストの生成を可能にする end-to-end なテキスト対話フレームワークであり，入力テキストから意味的文脈と感情的文脈を分散表現の抽出によって応答時の感情カテゴリの推定を行う感情制御部と，注意機構付き Seq2Seq に感情カテゴリラベルを付与することで感情的な対話応答生成を行う感情応答生成部の二つのニューラルネットワークによって構成される．推定された感情カテゴリに即して生成された感情的応答が適切なものであるのかを評価するために，対話生成指標を用いた自動評価とクラウドソーシングを使用して集めた 100 人の被験者による人手評価によって既存手法と提案手法の比較実験を行った結果，提案手法が生成精度や対話の自然さ，意味的整合性などの指標で上回ることを示した．

キーワード：感情制御，対話システム，アフェクティブコンピューティング

ER-Chat: Response Text Generation for Dialogue System with Emotional Regulation

Abstract: Expressing emotions in a dialogue system can improve users' life satisfaction and increase positive interactions. Recently, an emotional response generation method using neural networks has been proposed, which enables us to generate more natural and smooth interactive responses. However, the emotional response generation methods in the existing studies do not provide emotional regulation that determines the emotion of the response based on the emotional state of the interaction partner and the content of the speech, such as in human dialogue communication. In this research, we propose an ER-Chat that uses emotional regulation to estimate the emotional category of a response from the user's speech input text to generate an emotional dialogue response. This method consists of two neural networks: an emotion regulation part that estimates appropriate emotion categories in response to a response using a distributed representation of the semantic and emotional context of the input text, and an emotion response generator part that generates emotional text responses by assigning emotion category labels to Seq2Seq with Attention mechanism. It is an end-to-end framework that enables the generation of response text for appropriate emotion categories from text input by pre-training on a large dialogue pair dataset with emotion labels. To evaluate the appropriateness of the estimated affective categories and the generated affective responses, we compared existing and proposed methods by automatic evaluation using a dialogue generation index and manual evaluation with 100 subjects using crowdsourcing. The results showed that the proposed method outperformed in terms of accuracy, fluency and semantic consistency.

Keywords: Emotion Regulation, Dialogue System, Affective Computing

ⓒ 2020 Information Processing Society of Japan

Vol.2020-HCI-190 No.26
Vol.2020-UBI-68 No.26
2020/12/9

1. はじめに

Apple 社の「Siri」や Amazon 社の「Alexa」などに代表される対話システムは一般的に，ユーザの発話に応じて家電の操作や天候の確認などのタスクを行うことを目的としたタスク指向型と，特定のタスクは行わずユーザと雑談や会話などを行うことを目的とした非タスク指向型に分類される．中でも，非タスク指向型に分類される雑談対話システムは，特定のタスクを行うことによってユーザの手助けを行うことはないが，雑談を継続させることでユーザの話し相手になって楽しませたりおもてなしを行うことが可能である．また近年では，雑談対話システムでより円滑なコミュニケーションを実現するために感情を扱う研究が行われている．既存研究では，感情表現を行う対話システムとのインタラクションによってユーザの生活満足度の向上 [1] や肯定的なインタラクションの増加 [2] などの効果が認められているほか，ユーザの感情状態に合わせて対話のスタイルのルールを変更することによってより自然な対話 [3] が実現されているため，ユーザの感情を把握した上で感情的な応答を行うことが円滑なコミュニケーションにおいて非常に効果的である．また昨今の深層学習の飛躍的発展により，雑談対話システムの応答手法は従来のルールベース手法やマッチング手法ではなく，Seq2Seq[4] などのニューラルネットワークを用いた対話応答の生成手法が主流となっており，大規模コーパスを用いることにより柔軟な応答を可能にしている．さらには，ニューラルネットワークを用いて感情的な応答を生成する手法 [5], [6], [7] も提案されており，従来の生成手法で存在した汎用的な応答が多く生成されてしまう問題を解決し，より自然で感情的な応答生成手法を確立している．しかし，これらの感情的な対話応答生成の手法では，生成時に特定の感情カテゴリを付与することで，その感情に基づいた応答内容の生成に焦点が当てられている．一方で，人間同士の対話コミュニケーションでは，ユーザの感情状態や発話内容を考慮した上で適切な対応の決定をし，応答を行う感情制御を行っている．例えば，対話相手が悲しい感情状態である時に，ペットが亡くなって悲しんでいる相手に対しては共に悲しむ共感の応答を行うが，テレビゲームに負けて悲しんでいる相手に対しては楽しい感情で気持ちを切り替えさせるような応答を行うことが考えられる．そのため，対話システムがユーザの感情状態や発話内容を考慮する感情制御を行った上で対話応答を行うことは円滑でより人間らしい対話の実現に効果的であると考えられる．そこで本研究では，ユーザの発話入力テキストから感情的文脈と意味的文脈を考慮した感情制御を行い，感情的な内容で応答生成を行う ER-Chat を提案する．分散表現を用いて感情的文脈と意味的文脈を考慮し，応答時の感情カテゴリの推定と感情的な対話応答生成を行う二つのニューラルネットワークを用いることで，入力テキストに対して感情的内容で対話応答を行う end-to-end なテキスト対話フレームワークを実現する．対話生成における評価指標を用いた自動評価と，クラウドソーシングを用いた被験者 100 人による人手評価を行い，既存の感情的対話生成手法と比較して ER-Chat が対話の滑らかさや満足度，意味的整合性などの指標においてより優れていることを示した．

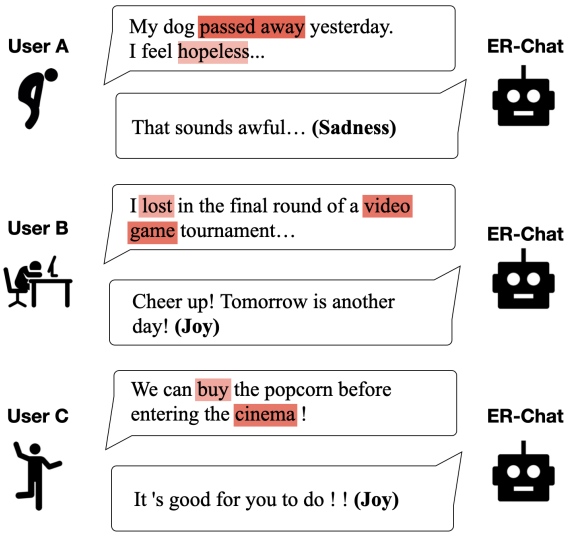

図 1 ER-Chat を用いた対話の入出力例

2. 関連研究

2.1 テキスト感情推定

感情を取り扱う技術として最も盛んに行われているのが，感情推定であり，古くから心理学や生物学的な観点からも数多く研究されている．感情状態は認知能力及び運動能力にも影響することが知られており，人間と機械間のコミュニケーションにも影響すると考えられている．コンピュータサイエンスの観点においても，人間の感情を推定し評価する研究が数多く行われており，Picard 氏によって Affective Computing[8] という研究分野が提唱されていることからも，コンピュータの発達に伴う感情推定の研究の重要性が伺える．自然言語処理における感情推定手法については，初期の研究は，テキストの極性を推定する研究 [9] が主だったが，その後，極性の度合いを求める研究 [10] や，

[1] 名古屋大学大学院工学研究科
Graduate School of Engineering, Nagoya University,Aichi 464-8603, Japan
[2] 慶應義塾大学大学院政策・メディア研究科
Graduate School of Media and Governance, Keio University, Fujisawa, Kanagawa, 252-0882, Japan
[3] 慶應義塾大学環境情報学部
Faculty of Information and Enviroment, Keio University, Fujisawa, Kanagawa 252–0882, Japan
[4] 名古屋大学未来社会創造機構
Intitutes of Innovation for Future Society, Nagoya University, Nagoya, Aichi, 464-8601, Japan
[a] shinsan@ucl.nuee.nagoya-u.ac.jp

ⓒ 2020 Information Processing Society of Japan

文章全体に極性を付与するのではなく，文章中の適切な範囲について極性を付与する研究 [11] がされてきた．また絵文字や Twitter のハッシュタグをメタ情報としてユーモアや皮肉，ヘイトを推定する研究 [12] なども行われている．テキストに発話者の感情があまり表現されないことと，テキスト内ユーザの受止め方は多種多様であることから，高精度で感情推定を行うことは難しいとされているが，近年ではニューラルネットワークを用いた感情推定手法も提案されており，BERT[13] などの技術を用いて感情推定を行う研究が盛んである．

2.2 対話応答生成

テキスト対話生成の手法は従来，手動で構築した特定のルールに基づいて応答を選択するルールベース手法が主流であった．しかし，莫大な量のルールを手動で記述する必要があるため非常にコストが高く，多くのトピックや発話文に対応することが困難であった．また，統計的抽出ベースの手法では，大量のテキストデータから現在の入力の応答として相応しいものを抽出するため，低コストで多くのトピックに対応可能であるが，生成される文章が毎回同じになってしまうため，機械的な応答になってしまう問題が存在した．そこで近年では，深層学習の発展による大規模な対話コーパスを用いて柔軟な対話の生成を低コストで行う生成ベースの手法が主流となっている．生成ベースの中でもっとも顕著であるのが 2014 年に Sutskever ら [4] によって提案された Sequence to Sequence(Seq2Seq) と呼ばれるニューラルネットワークである．入力テキストから出力テキストなど，時系列データを別の時系列データに変換することが可能なこの手法は対話文生成のほか，機械翻訳や画像からのキャプション生成などにも応用されている．

また，Seq2Seq を応用して，多様性を持った対話生成 [14] や個人に最適化された対話生成 [15] など，従来の機械的な対話手法を逸脱したより人間らしい対話生成が可能になっている．これらのように現在では，対話システムのほとんどがニューラルネットワークを用いた生成ベースの手法によって構築されている．

2.2.1 感情応答生成

1975 年に提案された Parry[16] はルールベースのアプローチながら感情を刺激できるメンタルモデルなどを用いるため，感情の発達に関与する最初の対話システムとされている．また，2014 年にはマイクロソフトが，ユーザーの感情的なニーズを認識できる共感的なソーシャルチャットボットである XiaoIce[17] を導入した例が挙げられる．さらには，高度な自然言語ベースの技術を活用した感情認識に基づく会話型メンタルヘルスケアサービス [18] や，ユーザの性格を学習することで共感的な反応を生成するインタラクティブシステム [19] なども提案されており，感情的対話システムの研究が盛んに行われている．ニューラルネットを用いた感情的な対話応答生成の研究においても多くの研究が行われており，Zhou ら [5] が提案した Emotional Chatting Machine（ECM) は，深層学習的アプローチを用いて大規模な感情表現を可能にした初めての対話システムである．また，感情埋め込み表現 [20], [21], [22] または強化学習を用いた手法 [23] などが提案されている．しかし，これらの研究の問題点として，感情的な応答の生成を行う際に，ユーザが最適な応答感情カテゴリを手動で選択しなければならない問題がある．そのため，本研究では対話相手の感情状態や発話内容を考慮した上で応答時の感情カテゴリを推定する感情制御を行い，感情的応答の生成を可能にする．

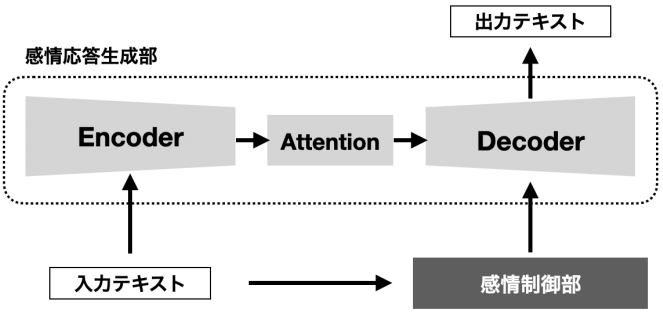

図 2 フレームワーク概要図

3. ER-Chat

本研究では，入力された発話テキストから応答時の感情カテゴリを推定し，その感情カテゴリに即した内容で対話応答を生成するテキスト対話フレームワークである ER-Chat を提案する．ER-Chat は応答時の感情カテゴリを推定する感情制御部と感情的な対話応答を生成する感情応答生成部の二つのネットワークから構成される．提案手法のフレームワーク概要図を図 2 に示す．

3.1 感情制御部

本ネットワークは，入力発話テキストから意味的文脈と感情的文脈の二種類を獲得し，これらの分散表現を用いて応答時の感情カテゴリを推定するネットワークである．本研究では，分散表現を獲得したのちに Self-Attention 手法を用いた Transformer[24] によるエンコードと全結合層によって，応答時の感情カテゴリラベルを正解データとして学習させ，推定を可能にする．

3.1.1 分散表現

テキスト内の各単語の表現を作成する最良の方法の 1 つとして，word2vec[25] などの分散表現がよく用いられる．分散表現の学習方法は，ほとんどの単語は同じ文脈で使用され，発生する単語が類似した意味を持つ傾向があるという言語学の分布仮説に基づいている．word2vec や単

語より小さな単位で埋め込みを行う同時に，文字レベルの N-gram(Character N-gram) である sub-word を用いる fastText[26] は似た意味を持つ単語を似ているベクトルで表現することを可能にしている．また，Agrawal ら [27] が構築した Emotional Embedding は感情に特化した分散表現であり，似た感情の単語を近いベクトルで表現する手法ことを可能にしている．本ネットワークでは，fastText を用いて意味的分散表現，Emotional Embedding を用いて感情的分散表現を獲得し，意味的分散表現と感情的分散表現の二種類を用いることで，発話テキストから応答時の感情カテゴリを適応的に推定に必要な表現力を拡張させる．

3.1.2 Transformer

Transformer を用いて単語の意味的分散表現と感情分散表現を用いることによって抽出した入力テキストの特徴をエンコードし，応答時の感情をベクトル形式で出力を行う．Transformer は再帰や畳み込みを用いず，Self-Attention 層と Feed Forward 層によって並列処理を行いながら学習を可能にする手法である．入力発話の $X = \{x_1, x_2, \ldots, x_T\}$ から獲得した意味的分散表現と感情的分散表現をそれぞれ Self-Attention 層と Feed Forward 層を用いてエンコードを行い，その出力を足し合わせて全結合層に送りソフトマックス関数によって one-hot ベクトルで表現された最適な感情ベクトル e の出力を行う．感情ベクトル e は，感情応答生成部の Attention 機構に渡すことにより，最適な応答感情カテゴリに即した感情的な応答生成を可能にする．

3.2 感情応答生成部

本ネットワークは，入力されたテキストデータなどの時系列データを対応する時系列データに変換する Encoder-Decoder から成る Seq2Seq フレームワークと注意機構 [28] によって構成される．Encoder では入力された発話文の単語系列を Long Short-Term Memory(LSTM) や Gated Recurrent Unit(GRU) などのリカレントニューラルネットワーク (RNN) に与え，発話文の情報を隠れ層のベクトルへ圧縮し，Decoder では隠れ層のベクトルの情報を基に単語を予測していき，応答文の生成を行う．$X = \{x_1, x_2, \ldots, x_T\}$ を入力すると時刻 t において式 (1) により隠れ層 h_t の状態を更新し，式 (2) により出力単語 y_t を計算する．

$$h_t = \mathrm{sigm}\left(W^{\mathrm{hx}} x_t + W^{\mathrm{hh}} h_{t-1}\right) \tag{1}$$

$$y_t = W^{yh} h_t \tag{2}$$

これを時刻 T まで繰り返すことによって，最終的な出力単語列 $Y = \{y_1, y_2, \ldots, y_T\}$ を得る．また，Decoder では，感情制御部にて出力された感情ベクトル e を注意機構と結合することによって感情的内容を含んだ応答 Y_e の生成を行う．Huang ら [6] の生成手法に基づいて，Encoder から受け取った隠れベクトル h と，感情適応部によって推定された感情ベクトル e を用いることで，感情的な応答を生成することが可能である．

この Encoder-Decoder は学習データ中の発話文に対して応答文の単語の予測確率が最大となるようにパラメータを更新することで，入力の発話文に対する最適な応答文を生成できるように学習される．

4. 評価実験設定

4.1 データセット

本システムで使用する対話データセットの要件として，発話分と応答分がペアになっている大規模対話データ，ならびにそれぞれに感情カテゴリラベルが付与されたデータである必要がある．そこで本研究では DailyDialog データセット [29] と，OpenSubtitles[*1]の字幕データセットを用いる．DailyDialog は，日常生活に関する様々なトピックにおける英文対話データセットであり，それぞれの対話文に感情カテゴリが手動でラベル付けされている．本研究で対象とする感情ラベルは Ekman[30] の基本感情として定義されている Anger，Disgust, Fear, Joy, Sadness, Surprise の 6 クラスとし，DailyDialog データセットの中から Ekman の基本感情に該当する対話分を抽出する．OpenSubtitles は，約 20,8000 件の映画やテレビの脚本から抽出した多言語字幕データセットであり，本研究では英文字幕データセットを用いる．OpenSubtitles の字幕データは感情ラベルが付与されていないため，4.1.2 にて感情カテゴリラベルの付与を行う．

4.1.1 前処理

収集した OpenSubtitle と DailyDialog の対話ペアデータに対して，3 単語以下 20 単語以上の文の除去，!,-. などの記号文字の除去，全て小文字に変換といった前処理を施した．

4.1.2 感情カテゴリラベルの付与

Opensubtitles の対話ペアデータには感情カテゴリラベルが付与されていないため，感情分類器の構築を行いそれらを用いて感情カテゴリのラベル付けを行う．感情分類器を構築する上での訓練データとして，EmotionX 2019[*2]で公開されている Friends データセットと EmotionPush データセットを用いる．Friends データセットは TV ドラマの Friends における発話から感情カテゴリラベルが付与された文章データセットであり，EmotionPush は匿名化された実際の Facebook Messenger チャットに感情カテゴリラベルが付与されたデータセットである．合計で 123,812 件の感情ラベル付きテキストが含まれている．本研究で対象とする Ekman[30] の基本感情における 6 種類の感情ラベルが付与されたデータのみを抽出し，訓練データとして感情分類器の学習を行う．感情分類器は，感情分類コンペ

[*1] http://www.opensubtitles.org/
[*2] https://sites.google.com/view/emotionx2019

ⓒ 2020 Information Processing Society of Japan

である EmotionX 2019 にて 79.5 の F1 スコアを達成した BERT モデル [31] の公開されている事前学習済みモデルとソースコードを用いて分類器の構築を行った.[*3]．以上の過程によって，OpenSubtitles と DailyDialog を合わせて 1,302,991 組の対話ペアデータを訓練データとして作成し，13,106 組の対話ペアデータをテストデータとして作成した．

4.2 実験パラメータ設定

本研究の提案手法である ER-Chat を構築するために，深層学習フレームワークである Pytorch を用いた．感情制御部と感情応答生成部は上記のデータセットを用いて事前学習を行う．単語の埋め込みサイズは 300, 語彙数は 25,000 とした．本実験を行う際の学習パラメータとして，感情制御部には 6 層と 4 ヘッドのマルチヘッド Transformer Encoder を持ち，各層の次元数は 2048，隠れ層は 100 次元である．感情応答生成部には Encoder と Decoder の各層に 256 次元の隠れ層を持つ 2 層 GRU 構造を持つ．学習プロセス全体で，パラメータの最適化関数には Adam[32] を採用し，学習率は 0.001，ドロップアウト率は 0.2 とした．また学習にはミニバッチ学習を採用し，ミニバッチサイズを 128 とした．エポック数は感情制御部が 30, 感情応答生成部が 10 とした．

4.3 ベースライン

ベースラインとして，注意機構付き Seq2Seq モデル (Att-Seq2Seq)[28] と，感情的な応答を可能にする Emotional Chatting Machine(ECM)[5] を採用する．ECM は手動で感情カテゴリの選択を行うため，本実験ではランダムな感情カテゴリで出力を行うように設定する．

4.4 評価指標

ER-Chat を用いて出力されるテキストの有効性を明らかにするために，本研究ではベースラインと比較した自動評価と人手評価を行う．

4.4.1 自動評価

自動評価では，入力テキストに対する ER-Chat の出力テキストの質を，ベースラインの出力テキストと比較することで評価を行う．文書生成モデルの全体的な評価として，翻訳などで BLEU スコアなどの評価指標が用いられるが，この指標は対話生成において人間の判断と相関がないことが明らかになっている [33]．そのため，本実験の応答生成精度評価には，正解文と出力文の類似性を算出する BERTScore[34] とモデルの出力の流暢さを評価する指標として用いられる Distinct-N(Dist-1, Dist-2)[14] によって評価を行う．BERTScore は生成文と正解文に含まれるトークンをそれぞれ BERT を用いることでベクトル表現へと変換し，これらの分散表現を用いて生成分と正解文の類似性を文脈全体から求めることが可能である評価指標である．Distinct-N は生成された応答に含まれる unigram と biggram の数を計算することで，多様性の度合いを算出することが可能である．これら二つの指標を用いて，提案手法における出力テキストの整合性や多様性を評価する．

4.4.2 人手評価

人手評価では，ER-Chat の対話応答の質を被験者のアノテーションによって評価を行う．テストデータに該当する DailyDialog データセットから 6 つの感情における入力テキストをそれぞれ 10 件ずつ，合計 60 件サンプル抽出する．DailyDialog データセットからテキストをサンプル抽出した理由として，日常生活の対話におけるテキストデータセットであるため被験者が対話を行うシーンを想定しやすいことと，手動で正確な感情ラベルが付与されていることから，適切な感情シナリオを提供できると考えたためである．被験者には，入力テキストに対して 2 つのベースラインと提案手法による 3 種類の出力テキストがランダムの順序で提示されるため，60 件の入力テキストに対して計 180 件の出力テキストが提示される．これら 180 件のテキストを，生成された応答における流暢さ (Fluency)，満足度 (Satisfaction)，意味的整合性 (Semantic Consistency)，感情の豊富さ (Emotinoal Eichness)，感情の適切さ (Emotional Appropriateness) の 5 つの指標から五段階のリッカート尺度で注釈をしてもらい，評価を行う．

5. 評価実験結果

表 1 に自動評価，並びに人手評価における実験結果を示す．また，人手評価に用いた DailyDialog データセットの各感情シナリオにおける対話応答例を表 6 に示す．

5.1 自動評価

ベースライン手法による出力テキストと比較して，ER-Chat における出力テキストは BERTScore, Distinct-1, Distinct-2 のそれぞれの値でベースライン手法を上回っていることが確認できた．BERTScore は生成された文章との類似性を評価する指標であるため，提案手法による感情制御を行うことで既存手法よりも感情で応答を可能にしていることが明らかになった．また, Distinct-1, Distinct-2 は生成された対話の多様性を評価する指標であるため，提案手法による評価が高くなっていることからより整合性のとれた多様な出力を可能にしていることが分かる．

5.2 人手評価

人手評価は, クラウドソーシングサービスの Amazon Me-

[*3] https://github.com/KisuYang/EmotionX-KU

表 1　被験者による実験で対象とした対話

Models	自動評価			人手評価 (平均値)				
	BERTScore (F 値)	Dist-1	Dist-2	Fluency	Satisfaction	Semantic Consistency	Emotional Richness	Emotional Appropriateness
Att-Seq2Seq	0.633	0.0272	0.099	3.131	2.307	2.564	2.537	2.497
ECM	0.671	0.0301	0.142	3.521	2.598	2.828	2.789	2.747
ER-Chat	**0.684**	**0.0332**	**0.165**	**3.733**	**3.124**	**3.284**	**3.217**	**3.274**

chanical Turk[*4]を用いて募集した 100 人の被験者によって行われた．そのうちインド在住の被験者が 40 人，アメリカ在住の被験者が 60 人であった．また，男性は 54 名，女性は 46 名で 20 代から 40 代以上の幅広い年齢層であった．表 1 に示す人手評価の値は，五段階のリッカート尺度によって注釈された平均値を示している．流暢さ (Fluency)，満足度 (Satisfaction)，意味的整合性 (Semantic Consistency)，感情の豊富さ (Emotinoal Eichness)，感情の適切さ (Emotional Appropriateness) の全ての指標において提案手法がベースライン手法を上回っていることがわかる．特に，満足度 (Satisfaction) においては Att-Seq2Seq よりも 0.817，ECM よりも 0.526 の値で上回っており，ER-Chat の出力テキストが比較的高い満足度を提供していることが分かる．また同様に感情の豊富さ (Emotinoal Eichness)，感情の適切さ (Emotional Appropriateness) といった二つの感情的指標もベースラインを上回っているため，提案手法による出力テキストが適切な感情を含んでいることが確認できた．

6. 考察

評価実験の結果から，以下の 2 点について考察を行った．

ダイバーシティ

人手評価を行うにあたり，各入力テキストに対して適切だと思う応答時の感情を 6 つの感情カテゴリから選択してもらった．その結果であるヒートマップを図 3 に示す．Joy シナリオにおいては，多くの被験者が Joy の応答を求めていることがわかり，Sadness シナリオにおいても同じく Sadness が求められていた．これらの感情シナリオにおいては，多くの場合被験者が対話相手に共感を求めていることがわかる．一方で，Anger や Disgust のシナリオにおいては，希望する感情カテゴリが分かれた．これらは，共感の Anger を所望する被験者もいれば気分転換を所望する被験者もいればまたそっとしておいて欲しい被験者もいるためだと考えられる．このように，応答時に被験者が希望する感情カテゴリは，同じ感情シナリオにおいてもかなり異なることが明らかになった．そのため，全てのユーザを満足させる対話システムを構築するためには，各ユーザの特性や選好性を学習してユーザごとにパーソナライズされる必要があると考える．

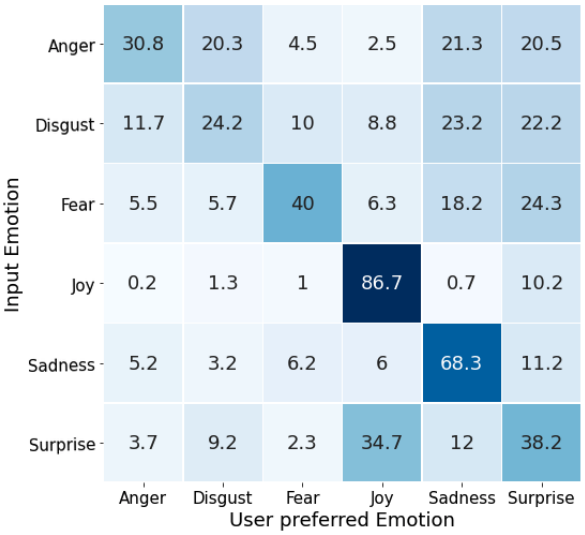

図 3　入力テキストに対する被験者が希望する応答時の感情カテゴリの選択率

ユーティリティ

本研究で対象にしていた対話は，テキストの入出力の 1 対 1 のターンで完結する対話であった．しかし，日常生活で行われる対話は複数のターンを横断して行われることがほとんどであり，一度の対話で完結するだけでなく複数のターンで対話を行い，また長期的にユーザとの過去の対話などを記憶しておくことで，より人間らしい対話システムの実現が可能になると考えられる．また，感情制御において適切な感情を推定する要素としては，本研究で取り扱った発話テキストの内容における意味的文脈と感情的文脈の他に，ユーザがどこにいて何をしているのかといったコンテクスト情報が必要である．家にいる場合と職場にいる場合，1 人でいる場合と友人といる場合などによってユーザが対話システムに求める感情的な応答は異なると考えられる．

7. おわりに

本研究では，ユーザの発話内容のテキストから分散表現を用いて意味的要素と感情的要素を抽出し，感情適応を行うことで感情情報を含んだ対話応答の生成を行うテキスト対話フレームワークである ER-Chat を提案した．Opensubtitles から取得した対話データと DailyDialog デー

[*4] https://www.mturk.com/

ⓒ 2020 Information Processing Society of Japan

表 2 DailyDialog[29] データセットの入力文に対する対話応答例

Input Text	Output Text	
He is lazy. I'll never ever work together with him! **(Anger)**	Att-Seq2Seq	you are not a coward
	ER-Chat	**it 's not that easy it is !? (Anger)**
He's endless talking about stocks almost bores me to death. **(Disgust)**	Att-Seq2Seq	i m going to kill him
	ER-Chat	**i don 't want anything to do with it (Disgust)**
I think someone is stalking me. **(Fear)**	Att-Seq2Seq	i m not a traitor so bad for you
	ER-Chat	**i can 't believe it ! i 'm telling you (Surprise)**
I'm going to get married next month. **(Joy)**	Att-Seq2Seq	i m going to get married
	ER-Chat	**it is like a great deal of life ! (Joy)**
I didn't perform as well as I should have in the interview. **(Sadness)**	Att-Seq2Seq	i m not going to hurt you
	ER-Chat	**would you like to have a drink ? (Joy)**
I can't believe he came to the office in jeans! **(Surprise)**	Att-Seq2Seq	i m going to get a job for the job
	ER-Chat	**that s not good, i m surprised he s doing it. (Surprise)**

タセットを用いて，感情カテゴリを推定する感情制御部と感情的な応答生成を行う感情応答生成部の二つのニューラルネットワークを構築した．対話応答生成における評価指標である BERTScore と Distinct-N を用いた自動評価と，クラウドソーシングサイトを用いて集めた 100 人の被験者による人手評価を行いベースラインの手法と比較した結果，自動評価と人手評価の両方においてベースラインの手法を上回る結果を示した．今後の展望としては，マルチターンの対話における感情制御や感情制御のパーソナライズ化を行うことで，より人間らしく円滑な対話の実現を目指す．

謝辞 本研究は，JST CREST JPMJCR1882 の支援を受けたものです．

参考文献

[1] Prendinger, H., Mori, J. and Ishizuka, M.: Using human physiology to evaluate subtle expressivity of a virtual quizmaster in a mathematical game, *International journal of human-computer studies*, Vol. 62, No. 2, pp. 231–245 (2005).
[2] Prendinger, H. and Ishizuka, M.: The empathic companion: A character-based interface that addresses users'affective states, *Applied Artificial Intelligence*, Vol. 19, No. 3-4, pp. 267–285 (2005).
[3] Polzin, T. S. and Waibel, A.: Emotion-sensitive human-computer interfaces, *ISCA tutorial and research workshop (ITRW) on speech and emotion* (2000).
[4] Sutskever, I., Vinyals, O. and Le, Q.: Sequence to sequence learning with neural networks, *Advances in NIPS* (2014).
[5] Zhou, H., Huang, M., Zhang, T., Zhu, X. and Liu, B.: Emotional Chatting Machine: Emotional Conversation Generation with Internal and External Memory, p. 9.
[6] Huang, C., Zaiane, O. R., Trabelsi, A. and Dziri, N.: Automatic dialogue generation with expressed emotions, *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pp. 49–54 (2018).
[7] Sun, X., Chen, X., Pei, Z. and Ren, F.: Emotional human machine conversation generation based on SeqGAN, *2018 First Asian Conference on Affective Computing and Intelligent Interaction (ACII Asia)*, IEEE, pp. 1–6 (2018).
[8] Picard, R. W.: *Affective computing* (2000).
[9] Turney, P. D.: Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews, *Proceedings of the 40th annual meeting on association for computational linguistics*, Association for Computational Linguistics, pp. 417–424 (2002).
[10] Pang, B. and Lee, L.: Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales, *Proceedings of the 43rd annual meeting on association for computational linguistics*, Association for Computational Linguistics, pp. 115–124 (2005).
[11] Wilson, T., Wiebe, J. and Hoffmann, P.: Recognizing contextual polarity in phrase-level sentiment analysis, *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing* (2005).
[12] Carvalho, P., Sarmento, L., Silva, M. J. and De Oliveira, E.: Clues for detecting irony in user-generated contents: oh...!! it's so easy;-, *Proceedings of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion*, ACM, pp. 53–56 (2009).
[13] Devlin, J., Chang, M.-W., Lee, K. and Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding, *arXiv preprint arXiv:1810.04805* (2018).
[14] Li, J., Galley, M., Brockett, C., Gao, J. and Dolan, B.: A diversity-promoting objective function for neural conversation models, *arXiv preprint arXiv:1510.03055* (2015).
[15] Li, J., Galley, M., Brockett, C., Spithourakis, G. P., Gao, J. and Dolan, B.: A persona-based neural conversation model, *arXiv preprint arXiv:1603.06155* (2016).
[16] Colby, K. M.: Artificial paranoia; a computer simulation of paranoid processes (1975).
[17] Zhou, L., Gao, J., Li, D. and Shum, H.: The Design and Implementation of XiaoIce, an Empathetic Social Chatbot, *CoRR*, Vol. abs/1812.08989 (online), available from ⟨http://arxiv.org/abs/1812.08989⟩ (2018).
[18] Dongkeon Lee, Kyo-Joong Oh and Ho-Jin Choi: The chatbot feels you - a counseling service using emotional response generation, *2017 IEEE International Confer-*

ⓒ 2020 Information Processing Society of Japan

ence on Big Data and Smart Computing (BigComp), Jeju Island, South Korea, IEEE, pp. 437–440 (online), DOI: 10.1109/BIGCOMP.2017.7881752 (2017).

[19] Siddique, F. B., Kampman, O., Yang, Y., Dey, A. and Fung, P.: Zara Returns: Improved Personality Induction and Adaptation by an Empathetic Virtual Agent, *Proceedings of ACL 2017, System Demonstrations*, Vancouver, Canada, Association for Computational Linguistics, pp. 121–126 (online), DOI: 10.18653/v1/P17-4021 (2017).

[20] Asghar, N., Poupart, P., Hoey, J., Jiang, X. and Mou, L.: Affective Neural Response Generation, *CoRR*, Vol. abs/1709.03968 (online), available from ⟨http://arxiv.org/abs/1709.03968⟩ (2017).

[21] Colombo, P., Witon, W., Modi, A., Kennedy, J. and Kapadia, M.: Affect-Driven Dialog Generation, *arXiv preprint arXiv:1904.02793* (2019).

[22] Shantala, R., Kyselov, G. and Kyselova, A.: Neural Dialogue System with Emotion Embeddings, *2018 IEEE First International Conference on System Analysis & Intelligent Computing (SAIC)*, IEEE, pp. 1–4 (2018).

[23] Li, J., Sun, X., Wei, X., Li, C. and Tao, J.: Reinforcement Learning Based Emotional Editing Constraint Conversation Generation, *CoRR*, Vol. abs/1904.08061 (online), available from ⟨http://arxiv.org/abs/1904.08061⟩ (2019).

[24] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł. and Polosukhin, I.: Attention is all you need, *Advances in neural information processing systems*, pp. 5998–6008 (2017).

[25] Mikolov, T., Chen, K., Corrado, G. and Dean, J.: Efficient estimation of word representations in vector space, *arXiv preprint arXiv:1301.3781* (2013).

[26] Bojanowski, P., Grave, E., Joulin, A. and Mikolov, T.: Enriching Word Vectors with Subword Information, *CoRR*, Vol. abs/1607.04606 (online), available from ⟨http://arxiv.org/abs/1607.04606⟩ (2016).

[27] Agrawal, A., An, A. and Papagelis, M.: Learning emotion-enriched word representations, *Proceedings of the 27th International Conference on Computational Linguistics*, pp. 950–961 (2018).

[28] Bahdanau, D., Cho, K. and Bengio, Y.: Neural machine translation by jointly learning to align and translate, *arXiv preprint arXiv:1409.0473* (2014).

[29] Li, Y., Su, H., Shen, X., Li, W., Cao, Z. and Niu, S.: DailyDialog: A Manually Labelled Multi-turn Dialogue Dataset, *CoRR*, Vol. abs/1710.03957 (online), available from ⟨http://arxiv.org/abs/1710.03957⟩ (2017).

[30] Ekman, P., Friesen, W. V., O'sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., Krause, R., LeCompte, W. A., Pitcairn, T., Ricci-Bitti, P. E. et al.: Universals and cultural differences in the judgments of facial expressions of emotion., *Journal of personality and social psychology*, Vol. 53, No. 4, p. 712 (1987).

[31] Yang, K., Lee, D., Whang, T., Lee, S. and Lim, H.: EmotionX-KU: BERT-Max based Contextual Emotion Classifier, *CoRR*, Vol. abs/1906.11565 (online), available from ⟨http://arxiv.org/abs/1906.11565⟩ (2019).

[32] Kingma, D. P. and Ba, J.: Adam: A method for stochastic optimization, *arXiv preprint arXiv:1412.6980* (2014).

[33] Liu, C.-W., Lowe, R., Serban, I., Noseworthy, M., Charlin, L. and Pineau, J.: How NOT To Evaluate Your Dialogue System: An Empirical Study of Unsupervised Evaluation Metrics for Dialogue Response Generation, *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, Austin, Texas, Association for Computational Linguistics, pp. 2122–2132 (online), DOI: 10.18653/v1/D16-1230 (2016).

[34] Zhang, T., Kishore, V., Wu, F., Weinberger, K. Q. and Artzi, Y.: Bertscore: Evaluating text generation with bert, *arXiv preprint arXiv:1904.09675* (2019).

ⓒ 2020 Information Processing Society of Japan