

日本語を対象とする感情を考慮した対話応答生成に関する研究

本間 心¹ 武田 敦志²

東北学院大学 教養学部情報科学科

¹s1657147@g.tohoku-gakuin.ac.jp, ²takeda@cs.tohoku-gakuin.ac.jp

1 はじめに

近年、ニューラルネットワークを用いて文章を分析することにより、その文章に対して適切な応答を自動で生成することが可能となった。特に、Recurrent Neural Network (RNN) である Sequence to Sequence (Seq2Seq)[6] を用いることで、高い精度の文章生成を実現している。また、Feed Forward Network (FNN) である Transformer を用いた文章生成方法も提案されている [7]。従来の RNN と比べ、Transformer は文章生成の精度が高いだけでなく、並列化することで訓練時間を削減できるという利点がある。これらの技術は、問い合わせやエンターテインメントにおける自動応答システムの開発に応用できると考えられる。円滑なコミュニケーションを行うためには、対話相手の感情を考慮する必要がある。Emotional Intelligence は心の知能指数 (EQ) と呼ばれ、自分や他人の感情を理解したり、感情を制御したりする能力である。これが高い人は社会において円滑にコミュニケーションを取ることができ、目標を達成しやすくなると考えられる。また、感情は人間の合理的な思考や意思決定にも必要な役割を果たしており、人間と自然かつ知的に交流するコンピューターには、感情を認識し表現する能力が必要とされている [3]。

感情を考慮した対話応答生成に関する研究として、研究は従来から行われており、英語の文書に対して Seq2Seq を用いることで感情カテゴリ（以降、感情と呼ぶ）を考慮した応答文章を生成する手法が提案されている [1, 2, 5]。また、中国語の文章に対しても、文章に含まれる感情表現に着目することにより、応答文章生成の精度を改善できることが示されている [8]。しかし、現在までに日本語の感情表現に着目した応答文章生成手法は報告されていない。本稿では、日本語の文章に含まれる感情表現を認識し、感情表現を考慮して日本語の応答文章を生成する手法を提案する。また、感情表現を考慮可能な新しい Transformer である

Emo-Transformer を述べ、Emo-Transformer が従来よりも高い精度で応答文章を生成できることを示す。

2 関連研究

機械翻訳のための RNN として Seq2Seq が提案された [6]。Seq2Seq では、入力文と出力文のペア (X, Y) 、 $X = x_1, x_2, \dots, x_m, Y = y_1, y_2, \dots, y_m$ に対して、確率 $P(Y|X)$ を最大化することで、任意の入力 X に対して推定された文 Y' を生成する。Encoder は入力文を、その特徴を表現するベクトル表現に変換し、Decoder に供給する。Decoder はベクトル表現と入力文から予測される出力文を生成する。Seq2Seq などの RNN を用いた手法では、文章の先頭から順番に処理していくため、並列化などの手法を用いて訓練時間を削減すること事は難しい。一方、自然言語処理を目的とした FNN である Transfmoer が提案されている [7]。Transformer も Seq2Seq と同様に、入力文をベクトル構造に変換する Encoder と、ベクトル表現と入力文から出力文を生成する Decoder で構成される。ただし、Seq2Seq と異なり、Transformer は FNN であるため、並列処理などの効率的な計算方法を導入することで訓練時間を削減することが可能である。また、Attention と呼ばれる機能を導入することにより、入力文に含まれる単語の前後関係を学習することが可能である。

Seq2Seq や Transformer は $P(Y|X)$ の確率を最適化することを目的とする。一方、文章と同時に感情を Seq2Seq の入力とすることで、感情を考慮した応答文章を生成する手法が報告されている [1, 2, 5]。これらの手法では、Seq2Seq の Encoder や Decoder に感情データを入力することにより、従来よりも高い精度で応答文章を生成することに成功している。一方、現在までに、日本語の感情表現を対象とした場合の応答文章生成手法や、感情を考慮した Transformer の適切な構成方法は報告されていない。

3 提案手法

本稿では, Transformer に感情を加えた, Emotion Transformer(Emo-Transformer) を提案する. Emo-Transformer は, どのように Transformer へ感情を加えるかによって, 2つの手法を提案する. 1つ目の手法は, Decoder の Positional Encoding(PE) 後に感情ベクトルを結合させる Emotional Encoding(EE) である. ここでは次式の計算を行う:

$$X_{enc} = \text{Concat}(P, EW^E) \quad (1)$$

ここで, P は単語の位置を表現した値であり, E は感情を表現した One-hot ベクトル, W^E は E をベクトル表現に変換する行列である. この式 (1) で計算した X_{enc} をデコーダの入力値に足し合わせる.

そして2つ目の手法は, Multi-Head Attention に感情を加える Emotion Query(EQ) である. この手法では各 Multi-Head Attention ブロックで次式の計算を行う:

$$\text{MultiHead}(Q, K, V, E) = \text{Concat}(\text{head}_1, \dots, \text{head}_h)W^O \quad (2)$$

where $\text{head}_i =$

$$\text{Attention}(QW_i^Q \odot EW_i^E, KW_i^K, VW_i^V) \quad (3)$$

ここで h は attention layer の数であり, Q, K, V , は文章を表現したベクトル, E は感情を表現した One-hot ベクトル, W_i^Q, W_i^K, W_i^V はそれぞれのベクトルを変換する行列である. Emotional Encoding と Emotion Query の計算手順を図 1 に示す.

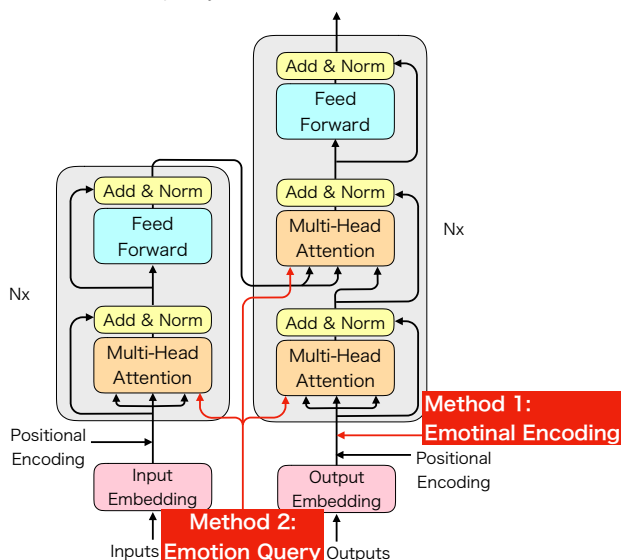


図 1: Emo-Transformer の構造図

表 1: テキストデータの検索に用いた単語

感情	単語
悲しい	亡く, 死ぬ, 死んだ, 悲し, 祖母, 祖父
不安	不安, 受験, 将来, 就職
怒り	怒り, 怒る, 理不尽, 裏切り
嫌悪感	嫌悪, 上司
信頼感	信頼, 友人
驚き	驚く, 結婚, 驚き, 驚いた
楽しい	旅行, 楽しい, 行く, 行った, 遊ぶ, 遊んだ

4 実験

4.1 感情付き対話データの作成

感情付きの対話データは, 既存のデータセットが無いため, 対話データに対して感情分類モデルを用いて, 感情付きの対話データを作成した. 感情分類モデルには, クラウドソーシングにより感情表現が収集されている, 日本語感情表現辞書: Japanese linguistic Inquiry and Word Count(以下, JIWC)[9]を使用した. 感情は Plutchik の感情の輪を参考に設定され, 悲しい, 不安, 怒り, 嫌悪感, 信頼感, 驚き, 楽しいの 7つの感情に分類されており, 関連する単語が各感情にカテゴライズされている. しかし, このままでは一部の文章に特化した感情分類モデルができる可能性があるため, 辞書から選出した代表的な単語を中心にテキストデータを収集し, 多種多様な文章を学習させた分類モデルを用いて, 感情分類モデルを作成した. 単語には, TF-IDF 値によって算出されたスコアが高い上位 6 単語を抜き出し, 重複するものを除外し, 検索に適するような形に変形した. テキストデータの収集には Twitter API¹ を使用し, 選出した単語が含まれているツイートを集めた. 最終的に使用した単語を表 1 に示す. データは各感情ごとに 178944 件収集し, そのうち約 10 % を検証用データに使用した. 分類モデルについては, BERT (Bidirectional Encoder Representations from Transformers) を用いた. BERT は, Transformer をベースとし, 大規模なテキストコーパスで事前学習を行い, 各タスクでの学習を行うことで, 感情分析を含む多くの自然言語処理のタスクで最先端の精度を達成している. また, 事前学習済みモデルを使用することで, 学習に用いるデータが少なくても高い精度を達成できる. 感情分類に際し, BERT の事前学習済みモデルの出力層に感情分類用の層を追加した. 事前学習済みモデルには, hottoSNS-bert: 大規模 SNS コーパスによる文分散表現モデル [4] を使用した. これは 8,500

¹<https://developer.twitter.com/en/docs>

表 2: 構築した対話データに関する統計情報

ユーザ数	3728526
ユニークユーザー数	2420364
会話の長さ (平均)	4.470457199
会話の長さ (最大)	7248

万のツイートで構成されたコーパスを用いた事前学習モデルである。表 1 に示した単語を [MASK] という記号に置き換えて、10 Epoch 学習させ、F1-score は 79 % に到達した。

対話データについて、Twitter から対話データを収集した。収集には Twitter API を使用し、特定のクエリを含むツイートの中から、応答であるツイートを抽出し、対話の始まりまで（以降、会話と呼ぶ）を収集した。Twitter 等の SNS では砕けた表現やネットスラングなどが多く、丁寧なドメインを対象としたかったため、クエリには丁寧語や「私」、敬称などをクエリにした。収集は日本語のツイートを対象として、リツイートと呼ばれる引用を示すものは除外した。また、ボットと呼ばれるユーザ名に「bot」という文字列が含まれるユーザが投稿した発言や、自分自身に返信している発言は除外している。文章の正規化では、発言中に URL があれば<url>という記号に置き換え、先頭にユーザ名が含まれる場合は、<mention>という記号に置き換えた。最後に NFKC 正規化、小文字化などの処理を行い、10 文字以上 190 文字以下の正規化された本文が重複しないツイートを対象とした。対象としたツイートは 9914368 件で、ツイートされた期間は 2009 年 10 月 11 日から 2019 年 11 月 14 日に渡っていた。統計情報を表 2 に示す。こうして収集した対話データに対して、応答文のみのデータを使用して、学習させた感情分類モデルに分類させ、質問文に対する応答文とその応答文の感情によって構成される、感情付きの対話データを作成した。

4.2 応答生成モデルの作成

提案手法との比較に用いるベースライン手法として、Shapala らの研究 [5] で提案された、感情を考慮した Seq2Seq モデルである、Seq2Seq with emotion embeddings と、Huang らの研究 [1] で提案された、Seq2Seq に Attention 機能を組み込んだモデルであり、応答文章に感情表現を考慮したモデルである Enc-att を用いた。また、感情を考慮することにより文章生成の精度が変化するか評価するために、Seq2Seq、Attention 機能を組み込んだ Seq2Seq、Transformer については、感情を考慮しない通常のモデルも用意した。Seq2Seq の

Encoder と Decoder で用いる RNN には、一層の Long Short-Term Memory (LSTM) を使用した。LSTM の隠れ層は 512 次元として、パラメータの最適化には、Adam を用いた。また、対話応答生成では質問文と応答文は非常に似ていることが予想されるため、Embedding 層は共通化した。さらに、Seq2Seq では、文の語順を逆にすることで学習の進みが早くなり、最終的な精度も向上するとされているため、応答文の語順を逆にして学習した。提案手法である、Emotional Encoding と Emotion Query の両方を評価する。Transformer を用いたモデルの設定は、NUM_LAYERS:6, D_MODEL:512, NUM_HEADS:8, UNITS:2048, DROPOUT は 0.1 とした。パラメータの最適化には、Adam を用いて、学習係数は [7] で提案された学習係数を用いた。

提案手法及びベースライン手法の全てのモデルに共通する設定は以下のとおりである。ミニバッチ学習を採用し、ミニバッチサイズを 512 とした。また、対話データのサブワードへの分割を行った。サブワード分割には SentencePiece を用いて、語彙サイズは 16000 とした。対話データは全体の約 90 % を学習に用いて、約 10 % を検証用に用いた。学習は 20 Epoch 学習させ、検証データで算出した Perplexity が最も低いものを評価に用いた。

4.3 評価

文章生成の精度の評価には、一般的にモデルの予測性能の良さを評価する指標として Perplexity を用いた。応答文章に感情表現が含まれる正確さ（感情の精度）の評価には、対話データの検証用データからランダムに選ばれた 1024 件の質問文に対して、応答生成モデルを用いて、各感情を指定して生成した応答を、感情分類モデルを用いて感情分類し、目的とする感情と一致するかどうかを調べた。したがって、応答文章に感情表現が含まれる正確さである感情の精度は、感情分類モデルによって推定される。

表 3 と表 4 に評価結果を示す。表 3 は文章生成の精度の評価で、感情を考慮した場合は文章生成の精度が向上することを示している。提案手法は、感情を考慮

表 3: 結果（文章生成の精度）

モデル	Perplexity
Seq2Seq	2.0580
Seq2Seq with emotion embeddings	2.0553
Seq2Seq with Attention	2.0285
Seq2Seq with Attention Enc-att	2.0171
Transformer	1.9125
Emo-Transformer (Emotional Encoding)	1.9078
Emo-Transformer (Emotion Query)	1.9081

表 4: 結果 (感情の精度)

感情	Emo-emb	Enc-att	Emo-Trans(EE)	Emo-Trans(EQ)
悲しい	9.96%	10.54%	38.86%	31.93%
不安	6.15%	17.87%	26.36%	14.25%
怒り	9.17%	7.32%	68.75%	70.21%
嫌悪感	6.44%	4.58%	2.24%	2.24%
信頼感	26.56%	20.41%	26.66%	19.92%
驚き	31.44%	12.40%	22.85%	13.67%
楽しい	50.29%	38.86%	39.0%	12.40%
平均	20.00%	15.99%	32.10%	23.51%

した既存手法や、通常の Transformer などよりも精度が高くなった。また、Emotional Encoding の精度は、Emotion Query の精度よりも高い結果となった。一方、表 4 は感情の精度を評価したものである。こちらに関しても、Emotional Encoding の精度は、Emotion Query よりも高い結果となった。提案手法は怒りの感情についての表現をよく学習できている傾向にある。嫌悪感についての全体的なスコアが低いのは、嫌悪感という感情についての文章の特徴を学習できなかった可能性がある。日本語の SNS 等における対話において、嫌悪感というのはあまりはっきりとは表現しない傾向にあり、実際に使用した対話データでは、嫌悪感は全体の 3 % だった。これが学習に影響した可能性が高い。これに関しては対話データの構築方法などを見直し、今後改善したい。最後に、提案手法を用いた対話システムの応答生成結果の一例を表 5 に示す。生成には Emo-Transformer (Emotional Encoding) を用いた。

5 おわりに

本稿では、Transformer に感情を加えた、Emo-Transformer を提案し、既存手法と比べて文章生成の精度や感情の精度が向上することを示した。また、評価ではニューラルネットワークを用いた対話応答生成において、感情を加えて学習させることによって文章生成の精度の向上を示し、感情を学習に加えることの重要性を示した。

表 5: 対話例

感情	対話例
-	風邪引いた
悲しい	私は、胃腸炎で、胃腸炎が痛いんです。
不安	風邪ひかないようにね。
怒り	お大事になさってください
嫌悪感	お大事になさってください。
信頼感	お一、それは良かったですね!
驚き	風邪ひかないようにね!。私も風邪ひかないように気をつける!
楽しい	

謝辞

本研究を推進するにあたり、日本語感情表現辞書: Japanese linguistic Inquiry and Word Count を提供して頂いた奈良先端科学技術大学院大学、荒牧英治先生、柴田大作氏に厚く御礼申し上げます。また、hottoSNS-BERT: 大規模日本語 SNS コーパスによる文分散表現モデルをご提供頂きました、株式会社ホットリンク開発本部 R&D 部に厚く御礼を申し上げ、感謝する次第です。

参考文献

- [1] Chenyang Huang and Osmar R Zaiane. Generating responses expressing emotion in an open-domain dialogue system. In *International Conference on Internet Science*, pp. 100–112. Springer, 2018.
- [2] Chenyang Huang, Osmar R Zaiane, Amine Trabelsi, and Nouha Dziri. Automatic dialogue generation with expressed emotions. In *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.
- [3] Rosalind W Picard. Affective computing. 1997.
- [4] Sakaki, Takeshi, Sakae Mizuki, Naoyuki Gunji. Bert pre-trained model trained on large-scale japanese social media corpus. 2019.
- [5] Roman Shantala, Gennadiv Kyselov, and Anna Kyselova. Neural dialogue system with emotion embeddings. In *2018 IEEE First International Conference on System Analysis & Intelligent Computing (SAIC)*, pp. 1–4. IEEE, 2018.
- [6] 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.
- [7] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in neural information processing systems*, pp. 5998–6008, 2017.
- [8] Hao Zhou, Minlie Huang, Tianyang Zhang, Xiaoyan Zhu, and Bing Liu. Emotional chatting machine: Emotional conversation generation with internal and external memory. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.
- [9] 柴田大作, 若宮翔子, 伊藤薫, 荒牧英治. Jiwc: クラウドソーシングによる日本語感情表現辞書の構築. 言語処理学会 第 23 回年次大会 発表論文集, pp. 771–774, 2017.