テキスト対話における受信者感情予測の個別化

菅野光 †1　　橋山智訓 †2
電気通信大学　　電気通信大学

1. はじめに

近年テキストを主体とした対話（以下テキスト対話）の機会は急増しており，テキスト対話における不安な経験やトラブルの発生が増加している．総務省情報通信政策研究所 [1] によると，LINE, X（旧 Twitter）の利用率はそれぞれ 2014 年の 55.1%, 21.9% から 2023 年の 94.9%, 49.0% へと大幅に増加している．バイドゥ株式会社と北海道函館西高等学校の共同調査 [2] では，テキスト対話によるトラブルや不安な経験の多さ，テキスト対話の支援機能への需要の高さが示されている．

テキスト対話におけるトラブルを防止する支援方法の一つには，メッセージの送信直前に送信内容に応じてフィードバックを与え，送信者の行動変容を促す方法が考えられる (図 1)．そのためには，送信内容からそれを読む受信者の感情を送信前に予測する（以下，受信者感情予測）必要がある．加藤らの実験 [3] では，受信者の感情の生じ方は受信者によって異なることが示されており，適切な支援の実現には受信者への個別化が重要といえる．例えば，相手をからかう送信内容に対して，受信者が冗談だと解釈してポジティブに感じやすい人であるか，真に受けて不快に感じやすい人であるかによって，支援内容を変えるべきである．受信者感情予測を試みた先行研究 [4][5] では受信者への個別化は考慮されていない．本研究は，受信者に対する受信者感情予測の個別化を目指す．

2. 関連研究

2.1. テキスト対話における受信者感情予測

長谷川ら [4] は，感情ラベル付きの対話コーパスを作成し，直前 2 発話から受信者の感情を予測するモデルを学習させ評価した．対話コーパスは X(当時 Twitter) から収集した投稿や返信を元に作成されている．テストデータに対する 8 感情予測の F1 値の平均は 0.567 であった．

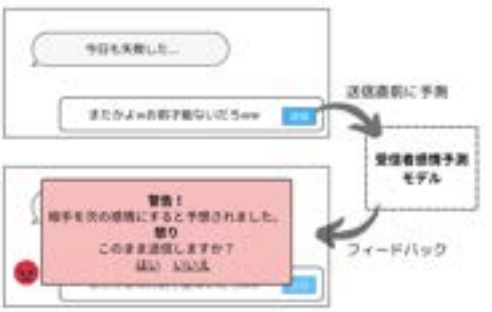

図 1　テキスト対話におけるトラブル防止支援システムのイメージ

古賀ら [5] は，テキスト対話における受信者の感情を推定するタスクを 4 種に分類し，それぞれに対応したモデルを学習させ評価した．これらのモデルは事前学習モデルを用いたファインチューニング手法で学習された．本研究が扱う受信者感情予測に該当するモデルでは，感情極性予測の F1 値の macro 平均は 0.55 程度であった．

2.2. 感情分析の個別化

鈴木ら [6] は書き手の感情分析を書き手に個別化できるモデルを提案した．このモデルは，同一の書き手の投稿を複数受け取ることで，書き手に特有の特徴ベクトルを得て推定を行う．この手法は，既存手法と異なり書き手の性格情報といった追加情報の付与が不要である上，既存手法を超える推定性能を示した．本研究ではこの手法を応用し個別化を試みる．

3. 目的

本研究では，テキスト対話における受信者感情予測について，受信者に応じた個別化手法を提案する．提案手法を評価するため，提案手法のモデルと比較用の単純なモデルで同じデータを学習させ，性能を比較する．

Personalization of Recipient's Emotion Recognition in Conversation

†1 HIKARU KANNO, The University of Electro-Communications
†2 TOMONORI HASHIYAMA, The University of Electro-Communications

4. 方法

4.1. 対話における受信者感情予測

本研究では，予測には受信者が読む送信内容とその直前 1 発話の内容を利用する．学習データは，3 発話分のテキストメッセージを元に，1 発話目と 2 発話目の送信内容及び 3 発話目に表現された感情情報で構成する．これらは，受信者が読む送信内容 (2 発話目)，その直前 1 発話の内容 (1 発話目)，ある送信内容を読む受信者の感情 (3 発話目に表現された感情) を表す．

予測は 8 感情から 1 つを選択することとする．8 感情は，Plutchik の感情の輪 [7] を元にした「嬉しい，悲しい，期待，驚き，怒り，恐れ，嫌悪，信頼」とする．

予測は深層学習モデルにより実現する．モデルの構築には事前学習済みの BERT[8] を利用し，ファインチューニング手法によって学習させる．モデルへの入力は，1 発話目と 2 発話目の送信内容 (文章) をトークン化して特殊トークン ([SEP]) で結合したトークン列とする．モデルは入力されたトークン列を特徴ベクトルに変換し，それを線形変換することで予測結果を出力する．8 感情に対する分類問題として学習させることで，予測モデルを実現できる．本研究では，事前学習済みの BERT として日本語の Wikipedia の文章を学習した bert-base-japanese-whole-word-masking[*1]を使用する．

4.2. 提案手法

本研究の提案手法は，鈴木らの提案手法 [6] を応用したものである．鈴木らの手法では同一の書き手の文章をモデルに複数入力するが，提案手法では同一の受信者の対話をモデルに複数入力する．

提案手法におけるモデルの構造と順伝播の模式図を図 2 に示す．このモデルの構造は，特徴量抽出を行う事前学習済みモデルと線形変換を行う全結合層の間に Self-Attention 層を加えた構造である．Self-Attention 層とは，各入力に対し他の入力との関連性を計算し，重要な要素を強調する機構である．モデルに与えられた各対話は事前学習済みモデルによって特徴ベクトルに変換され，Self-Attention 層に入力される．Self-Attention 層での変換は同時に入力される他の入力によって決まる．すなわち Self-Attention 層に入力された特徴ベクトルは，受信者ごとに異なる変換がされる．この変換で得た特徴ベクトルを使って予測することで，受信者ごとに異なる予測ができる．

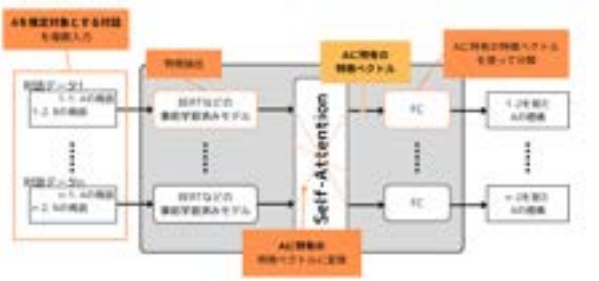

図 2　提案手法におけるモデルの構造と順伝播の模式図

5. データセットの作成

本研究でのモデルの学習には，2 人の人間同士のテキスト対話を記録したデータが必要である．特に個別化手法の評価をする観点から，対話者が識別できること，多様な対話者を含むことが必要である．一般公開されたデータセットにはこれらの条件を全て満たすものがない．そこで本研究では独自にデータセットを作成した．

データの作成には，Bluesky[*2]上の投稿と返信を収集した．その結果，642 人の受信者に関する 10534 通りの対話を得た．収集したデータに対して感情ラベルのアノテーションを実施した．アノテーションには感情分析モデル luke-japanese-large-sentiment-analysis-wrime[*3]を利用した．このモデルには各データの 3 発話目を入力し，得られた出力を Softmax 変換で確率分布に変換したものを感情ラベルとした．

感情分析モデルによるアノテーション結果を評価するため，収集したデータから無作為抽出した 100 通りの対話に対して，5 人の作業者によるアノテーションを実施した．作業者によるアノテーションとの比較の結果，モデルによるアノテーションで作成したデータは本研究の学習データに使用可能と判断した．

6. 実験

提案手法が個別化を実現できるかを検証するため，図 2 に示す提案手法のモデルと，Self-Attention 層を持たない比較用の単純なモデルでそれぞれ学習させ，性能比較を実施した．

学習データは 5 章で作成したデータセットを無作為に 6:2:2 の割合で訓練, 検証, テスト用のデータに分割し，長さ 512 のトークン列に符号化したものを使用した．学習の設定を表 1 に示す．

[*1] https://huggingface.co/tohoku-nlp/bert-base-japanese-whole-word-masking

[*2] https://bsky.app

[*3] https://huggingface.co/Mizuiro-sakura/luke-japanese-large-sentiment-analysis-wrime

表 1　各モデルの学習の設定

設定	比較モデル	提案モデル
loss func.	Weighted Cross Entropy	
batch size	32	4[*4]
optimizer	AdamW	
weight decay	0.1	0.1
warmup steps	1epoch	1.5epoch
lr	3e-5	3e-5
max epochs	10	15
dropout	0.1	0.1

7. 結果と考察

比較モデルの学習は 6 エポックで早期終了し，提案モデルの学習は 15 エポックで終了した．検証データへの損失は，比較モデルでは 3 エポック目，提案モデルでは 7 エポック目が最小となった．テストデータによる各モデルの性能評価結果を表 2 に示す．

表 2　評価結果

指標	比較モデル	提案モデル
accuracy	0.54	**0.55**
macro-F1	**0.46**	0.43
weighted-F1	0.54	**0.55**

表 2 を見ると，提案手法によるモデルの性能は，比較用の単純なモデルの性能と大きな差が見られない．したがって，本研究の提案手法は今回の実験では個別化を実現できなかったと結論づけた．

個別化が実現できなかった原因として，受信者 1 人あたりのデータ数が少なかったことが挙げられる．鈴木ら [6] は書き手 1 人あたりの投稿数が 600 件以上となる前処理をしていたが，今回使用したデータでは受信者 1 人あたりのデータ数は平均 16.4 件であった．

学習の様子として，バッチサイズが小さいほど学習が不安定になり，学習後の性能がバッチサイズに大きく依存していた．これは，今回の学習データが不均衡であったことが原因と考えられる．今回は損失関数の重みを調整することで学習の安定化を図ったが，それだけでは不十分であったと考える．

[*4] 提案モデルの学習には 8 通りの対話のデータを 1 単位にまとめたデータを使用している．バッチサイズ 4 のミニバッチは 32 通りの対話のデータを持つ．

8. 今後の課題

8.1. 受信者群に対する個別化

個別化実現のための改善案として，受信者をクラスタリングし，似た特徴を持つ受信者群への個別化を試みるべきだと考える．受信者 1 人あたりのデータが少なくても受信者 1 群あたりのデータは十分に確保できる．また加藤らの実験 [3] では，受信者の感情の生じ方は受信者の感情特性により分類した 3 群間で異なることが示されている．この実験における感情特性とは，特定の感情経験の頻度または特定の感情反応の閾値の低さに関する，長期的に安定した個人の傾向または特性を指す．感情特性により分類した受信者群に対して提案手法を適用することで，個別化の実現が期待できる．

感情特性により受信者を分類するには，SNS 上の過去の投稿内容を感情分析するつもりである．そうすることで例えば，喜びを表現した投稿が多い群，怒りを表現した投稿が多い群，などと受信者を分類できる．今回は同一の受信者の対話で入力を構成したが，今後は同一の受信者群の対話で入力を構成し個別化を試みたい．

8.2. 感情ラベルの形式変更

感情ラベルの形式に関して，モデルの性能向上に向けた改善案を以下に列挙する．

- いずれの感情にも属さない中間状態を追加する
- 似た感情同士が隣合うような間隔尺度で表現する
- 複数選択を許容する（マルチラベル分類）

特定の感情に属さない中間状態を追加することで性能向上が期待される．古賀らの研究 [5] では「positive, negative, neutral」の 3 分類で学習させた．3 分類の macro-F1 と neutral を除く 2 分類の macro-F1(macro-F1 without neutral) を比較すると 3 分類の macro-F1 値の方が高い結果が示された．本研究と同条件では，macro-F1 が約 54%，macro-F1 without neutral が約 38% であった．

本研究のテスト結果の評価では，モデル出力と正解ラベルが一致すれば正解，不一致ならば不正解としていた．しかし実用上は，不正解だが正解と似た感情を答える場合 (「嬉しい」に対し「期待」など) はさほど問題でなく，全く異なる感情を答える場合 (「嬉しい」に対し「怒り」など) は大きな問題となる．似た感情同士が隣合うような間隔尺度で感情ラベルを表現すれば，損失関数に正解との間隔に応じた重みを付けられる．そうした重みを持つ損失関数を学習に用いれば，たとえ不一致でもなるべく似た感情を予測するようにモデルを訓練できる．

本研究では 8 感情から 1 つのみ選択していた (シングルラベル分類) が，複数の感情状態の選択を許容すること (マル

チラベル分類) でより適切な予測ができると考える．実際の受信者の感情は常に 1 つとは限らず複数の感情を同時に抱くこともあり得る．そうした場合に単一選択の予測で正解するのは難しく，複数選択の予測によって正解や部分的な正解が容易になる．

8.3. 不均衡データへの対応

不均衡な感情ラベルを持つデータの学習を安定化させるために本研究では損失関数の重みを調整したが，それだけでは不十分であった．アンダーサンプリングやオーバーサンプリングによりデータの不均衡さを是正したり，異常検知問題としての学習を試してみたい．小さなバッチサイズでも安定した学習が可能になれば，少ない計算量での学習が可能となる上，性能向上が期待できる．

8.4. 事前学習済みモデルの変更

本研究で用いた BERT(bert-base-japanese-whole-word-masking) は Wikipedia の文章で事前学習されたモデルである．SNS 上の投稿を事前学習したモデルを使えば，くだけた表現に対しても適切に特徴抽出でき，高い性能を発揮する可能性がある．SNS 投稿のテキストを事前学習したモデルには，hottoSNS-BERT[9] や JTweetRoBERTa[10] などがある．

対話における感情分析を想定した事前学習済みモデルを使うことも有効だと考える．例えば BERT-ERC[11] などは発話間の関係について特徴抽出できるなどの特徴がある．こうしたモデルにより対話の文脈情報まで捉えることで性能向上が期待できる．

8.5. 実用上の性能評価

本研究はモデルの性能をテストデータによって評価したが，テストデータに対する性能が，実用における汎化的な性能を示すとは限らない．今回のテストデータは Bluesky 上で収集したデータに限定されており，Bluesky 以外での対話に対する性能は示せない．また今回のテストデータは感情分析モデルによるアノテーションで作られたもので，実際の受信者の感情とは異なる．

実用的な汎化性能の測定するには，受信者本人の感情が付与されたテストデータが必要である．そうしたデータは，対話の各時点での感情を記録してもらったデータや自身の過去の対話にアノテーションしてもらったデータを収集すれば用意できる．しかしデータの偏りを減らしつつ効率良く大量のデータを収集するには，自動的に収集できる方法の検討が課題となる．

9. まとめ

本研究は，テキスト対話における受信者感情予測の個別化手法を提案した．独自にデータセットを作成し提案手法を評価した結果，提案手法は今回の実験では個別化を実現できなかった．個別化実現のためには受信者をクラスタリングし受信者群への個別化を試みる等の工夫が必要である．

参考文献

[1] 総務省：令和 5 年度情報通信メディアの利用時間と情報行動に関する調査報告書の公表 (2024). https://www.soumu.go.jp/menu_news/s-news/01iicp01_02000122.html.

[2] バイドゥ株式会社：北海道函館西高等学校探究チーム「ぶなしめじ君の冒険」と実施のテキストコミュニケーションの課題に対処するための調査報告 (2024). https://prtimes.jp/main/html/rd/p/000000821.000006410.html.

[3] 加藤由樹，加藤尚吾，杉村和枝，赤堀侃司：テキストコミュニケーションにおける受信者の感情面に及ぼす感情特性の影響： 電子メールを用いた実験による検討，日本教育工学会論文誌， Vol. 31, No. 4, pp. 403–414（オンライン），10.15077/jjet.KJ00004964312 (2008).

[4] 長谷川貴之, 鍛治伸裕, 吉永直樹, 豊田正史：オンライン上の対話における聞き手の感情の予測と喚起，人工知能学会論文誌, Vol. 29, No. 1, pp. 90–99（オンライン), 10.1527/tjsai.29.90 (2014).

[5] 古賀友里愛，神藤駿介，宮尾祐介：対話における発話に反映されない聞き手の感情推定，人工知能学会全国大会論文集， Vol. JSAI2024, pp. 4Xin246–4Xin246（オンライン），10.11517/pjsai.JSAI2024.0$_4$*Xin*246(2024).

[6] 鈴木陽也，山内洋輝，梶原智之，二宮　崇，早志英朗，中島悠太，長原　一：書き手の複数投稿を用いた感情分析，人工知能学会全国大会論文集，Vol. JSAI2024, pp. 3Xin2104–3Xin2104（オンライン），10.11517/pjsai.JSAI2024.0$_3$*Xin*2104(2024).

[7] PLUTCHIK, R.: Chapter 1 - A GENERAL PSYCHOEVOLUTIONARY THEORY OF EMOTION, *Theories of Emotion* (Plutchik, R. and Kellerman, H., eds.), Academic Press, pp. 3–33 (1980).

[8] Devlin, J., Chang, M.-W., Lee, K. and Toutanova, K.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding (2019).

[9] Sakaki, Takeshi, Mizuki, S., Gunji, N.：BERT Pre-trained model Trained on Large-scale Japanese Social Media Corpus, https://github.com/hottolink/hottoSNS-bert (2019).

[10] 高須遼, 狩野芳伸：JTweetRoBERTa: 大規模 SNS 投稿テキストによる事前学習と各種タスクによる性能検証，言語処理学会 第 30 回年次大会 発表論文集，言語処理学会 (2024).

[11] Qin, X., Wu, Z., Cui, J., Zhang, T., Li, Y., Luan, J., Wang, B. and Wang, L.: BERT-ERC: Fine-tuning BERT is Enough for Emotion Recognition in Conversation (2023).