感情の顕現性を考慮した書き手の感情強度推定

岡留 司真　　白井 清昭
北陸先端科学技術大学院大学 先端科学技術研究科
{s2210038,kshirai}@jaist.ac.jp

概要

本論文では，書き手の感情の強さを推定する新しい手法を提案する．書き手の感情がテキストの表層にどれだけ明示的に表現されているかを感情顕現性と定義し，書き手と読み手の感情強度が付与されたデータセットからこれを推定するモデルを学習する．また，書き手の特徴を表す単純な特徴量として書き手 ID を導入する．未知の書き手に対して，それと類似した書き手を訓練データから選択し，その書き手 ID を入力として与える．感情顕現性と書き手 ID を感情強度推定モデルに組み込むことの有効性を実験により検証する．

1　はじめに

テキストの感情推定は，テキストに表出している感情を推定するタスクであり，様々な場面に応用されている．感情推定には，テキストの書き手もしくは読み手の感情を推定する 2 つのタスクがある．書き手の感情が必ずしも読み手に伝わらないことがあるように，書き手と読み手の感情は一般に異なる．また，書き手の感情はテキストの表層上の表現に反映されないこともあるため，書き手の感情推定は読み手の感情推定より難しいとされている [1]．

本研究では，与えられたテキストに対し，その書き手の感情の強さを推定することを目的とし，書き手の感情がどれだけ直接的にテキストの表層に現われているかを考慮した手法を提案する．書き手と読み手の感情が乖離しているときには書き手の感情は暗黙的に表現されているとみなし，両者の感情が乖離しているかを判定するモデルを学習する．さらに，これから得られるテキストの特徴を感情強度推定モデルに組み込む．また，書き手を識別する「書き手 ID」の情報も利用する．書き手が訓練データに出現しないとき，訓練データにおける書き手の中で最も類似した書き手を選び，その書き手 ID を用いる．書き手が既知・未知であるという 2 つの実験設定で評価実験を行い，提案手法の有効性を検証する．

2　関連研究

テキストに書き手や読み手の感情を付与したデータセットはいくつか公開されている．Buechel と Hahn は，Valance，Arousal，Dominance を感情クラスとし，それぞれについて書き手と読み手の感情強度を付与した EmoBank を構築した [2]．ただし，書き手の感情強度は書き手以外の第三者が推測して付与している．書き手よりも読み手の感情強度の方がアノテータ間の一致率が高かったと報告しているが，これは書き手の感情強度推定が難しいことを示唆する．Kajiwara らは，テキストに対して書き手の感情 (主観的な感情) と読み手の感情 (客観的な感情) の強度を付与したデータセット WRIME を構築した [3]．EmoBank とは異なり，書き手の感情強度はテキストの書き手自身によって付与されている．WRIME を元に感情強度を推定するモデルを学習した実験では，書き手・読み手両方の感情強度推定タスクにおいて，書き手の感情強度ラベルを用いて学習したモデルの方が読み手の感情強度ラベルを用いて学習したモデルよりも性能が高かったと報告している．

最近の感情強度推定に関する研究では，Bidirectional Encoder Representations from Transformers (BERT)[4] などの大規模言語モデルを利用する手法が多い．Alhuzali と Ananiadou は BERT に基づくマルチラベルの感情分析手法を提案した [5]．正解ラベルと不正解ラベルに対する予測スコアの差が大きくなるように損失関数を設定し，感情間の関係を損失関数に統合することの利点を示した．鈴木らは書き手の性格を考慮して書き手の感情強度を推定する手法を提案した [6]．Big-5[7] によって定義される書き手の性格の特徴と，BERT によって得られるテキストの特徴ベクトルを組み合わせることで，感情強

度推定の性能を向上させた．

本研究では書き手の感情強度推定に有効な特徴量を探究する．鈴木らの手法 [6] では書き手の性格を利用しているが，事前に性格診断を受ける必要がある．本研究では，感情強度が付与されたデータセットのみを利用することを想定し，書き手と読み手の感情がどれだけ乖離しているかの情報と，書き手の識別情報を利用する．

3 提案手法

3.1 問題設定

本論文では書き手の感情強度を推定するモデルを学習するためのデータセットとして WRIME[3] を用いる．WRIME は，ソーシャルメディアに投稿されたおよそ 35,000 件のテキストに対し，書き手と読み手の感情強度ラベルを付与したデータセットである．Plutchik の感情モデル [8] にしたがい，8 つの感情クラス (Joy, Sadness, Anticipation, Surprise, Anger, Fear, Disgust, Trust) のそれぞれについて，0, 1, 2, 3 のいずれかの感情強度のラベルが付与されている．0 はその感情を持たない (中立) であることを示す．書き手の感情強度はテキストを書いた人が付与し，読み手の感情強度は 3 名のクラウドワーカーによって付与されたラベルの平均値が付与されている．また，各テキストの書き手を識別する書き手 ID も付与されている．

本研究では，WRIME の仕様に沿ってタスクを設定する．与えられたテキストに対し，8 つの感情クラスのそれぞれについて，その書き手の感情強度の強さを 0, 1, 2, 3 のいずれかに分類する．ここではタスクを回帰ではなく分類問題と定義する．

さらに，分類対象のテキストの書き手が訓練データに出現するか否かによって 2 つのタスクを定義する．“Closed Task” では，テストデータのテキストの書き手が訓練データに出現し，同じ書き手のラベル付きデータをモデルの学習に利用できると仮定する．“Open Task” では，テストデータの書き手は未知である，すなわち同じ書き手のデータは訓練データに存在しないと仮定する．Open Task の方がより現実的な問題設定と言える．

3.2 感情強度推定モデルの学習

提案手法の感情強度推定モデルの概要を図 1 に示す．以下，その詳細を説明する．

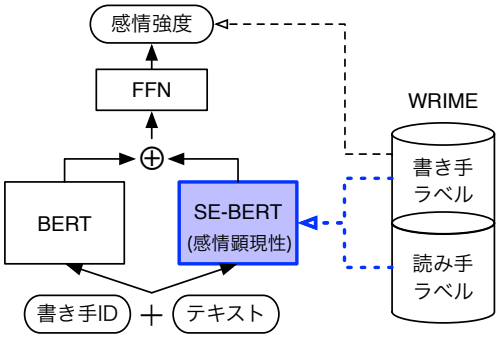

図 1 感情強度推定モデル

3.2.1 感情顕現性

書き手が暗に感情を示す場合など，書き手の感情がテキストの表層に現れないことがある．このとき，テキストの表層に現れる感情と実際の書き手の感情に乖離があり，このことが書き手の感情強度推定を難しくしていると考えられる．そこで，書き手の感情がどの程度テキストに直接的に表されているかの指標として感情顕現性 (Saliece of Emotion; SE) を導入する．感情語などによって書き手が直接的に感情を表現しているときには感情顕現性が高く，書き手が間接的に感情を表わしているときには感情顕現性が低いとする．

我々は，テキストにどれだけ書き手の感情が表出しているかは，テキストの読み手がその感情をどれだけ読み取れるかで測れると考える．この考えに基づき，感情顕現性を書き手と読み手の感情強度の違いと定義する．書き手と読み手の感情強度が一致していれば，感情顕現性は高いとする．

まず，テキストを入力とし，感情顕現性を予測するモデルを学習する．感情顕現性のラベルを，書き手と読み手の感情強度が一致しているときには 0，それ以外は 1 と定義する．[1] WRIME から感情顕現性ラベルが付与されたデータセットを作成し，これを用いて BERT をファインチューニングすることで感情顕現性を予測するモデルを学習する．以下，このモデルを SE-BERT と呼ぶ．

図 1 に示すように，テキストの書き手の感情強度を推定する際には，BERT から得られるテキストの埋め込み ([CLS] トークンに対する埋め込み) と，SE-BERT から得られるテキストの埋め込みを連結し，Feedforward Network (FFN) への入力とする．FFN は感情強度の各クラスに対する予測確率を出力

1) 読み手の感情強度は 3 名のクラウドワーカーの評価値の平均値が付与されているが，本実験で使用するときは四捨五入して整数に変換する．

する．書き手の感情強度ラベルを正解データとし，クロスエントロピーを損失関数として，BERT と FFN のパラメータを学習する．ただし，SE-BERT は事前に学習しておき，そのパラメタは更新しない．

3.2.2 書き手 ID

テキストに書き手の感情がどれだけ表出するかは，人によってバラツキがある．すなわち，直接的に感情を表す人もいれば，直接的な感情表現を避ける人もいる．したがって，書き手の情報を感情強度推定モデルに組み込むことで，その性能を向上させることができると考えられる．

本研究では，書き手に関する情報を組み込む最も単純な手法として，書き手 ID を入力として与える．具体的には，以下の 2 つの形式で BERT への入力に書き手 ID を加える．

P1 [CLS] [W-ID] 文

P2 [CLS] 文 [SEP] [W-ID]

ここで [W-ID] は特殊トークンで表された書き手 ID である．予備実験の結果，**P1** と **P2** とで感情強度推定の性能はほとんど変わらなかった．4 節の実験では書き手 ID は **P1** の形式で与える．

Closed Task では，モデルの学習時もテスト時も，書き手 ID は WRIME に付与されたものをそのまま用いる．Open Task では，モデルの学習時の書き手 ID は WRIME に付与されたものを用いる．テスト時は，書き手は未知であるため，以下の手続きで書き手 ID を決める．テストデータで同じ書き手によって書かれたテキストを事前学習された BERT を用いて埋め込み表現に変換し，その平均ベクトルを求める．訓練データにおける書き手についても同様に，その書き手によって書かれたテキストの埋め込みの平均ベクトルを求める．両者のコサイン類似度によって書き手間の類似度を計算し，訓練データの中で最も類似度の高い書き手の ID を入力として与える．ここでは，感情強度を推定したいテキストを書いた人のラベルなしテキストがある程度蓄積されていることを仮定している．

4 評価実験

4.1 実験設定

評価実験では，WRIME を訓練データ・開発データ・テストデータに 8:1:1 の割合で分割して使用する．Closed Task では，同じ書き手のテキストが訓練・開発・テストデータの全てに含まれるように分割する．Open Task では，まず書き手を 8:1:1 に分割し，それぞれの書き手が書いたテキストの集合を訓練・開発・テストデータとする．実験データの統計を表 1 に示す．

表 1 実験データの詳細

	Closed Task			Open Task		
	訓練	開発	テスト	訓練	開発	テスト
書き手	60	60	60	48	6	6
テキスト	28,000	3,500	3,500	28,000	3,500	3,500

実験では，8 つの感情クラスについて独立にモデルを学習し，感情強度推定の性能を評価する．以下の 4 つの手法を比較する．

BL テキストを入力し，BERT によって感情強度を推定するモデル．これをベースラインとする．

M-SE 感情顕現性を判定するモデルから得られる埋め込みを使用するモデル．

M-WID 書き手 ID を入力に与えるモデル．

M-SE-WID 感情顕現性判定モデル，書き手 ID の両方を用いるモデル．本研究の提案手法．

事前学習済み BERT として，東北大が公開している BERT base Japanese[9] を使用した．

4.2 感情顕現性判定モデルの評価

感情顕現性を判定するモデルの精度 (Precision)，再現率 (Recall)，F 値 (F-measure)，正解率 (Accuracy) を表 2 に示す．正解率はどの感情クラスについても 0.7 を超えているが，F 値は 0.63 から 0.75 程度である．F 値が正解率と比べて低いのは，書き手も読み手もその感情強度が 0(=中立) の場合，すなわち感情顕現性クラスが 0 の場合が多く，不均衡なデータに対して正解率が高く見積られているためである．

表 2 感情顕現性判定モデルの評価

	Precision	Recall	F-measure	Accuracy
Joy	0.77	0.74	0.75	0.77
Sadness	0.72	0.70	0.71	0.73
Anticipation	0.70	0.68	0.69	0.72
Surprise	0.67	0.64	0.65	0.74
Anger	0.64	0.62	0.63	0.85
Fear	0.66	0.63	0.64	0.81
Disgust	0.67	0.66	0.66	0.78
Trust	0.65	0.64	0.65	0.78

4.3 感情強度推定モデルの評価

Closed Task における感情強度推定モデルの精度，再現率，F 値 (感情強度クラス 0,1,2,3 に対するマクロ平均)，ならびに正解率を表 3 に示す．太字は 4 つの手法の中で最良の値を示している．F 値と正解率に着目すると，大部分の感情クラスについて，提案手法 (M-SE-WID) によって一番高い評価値が得られている．このことから，感情顕現性を感情強度推定モデルに組み込むこと，話者 ID を利用することの有効性が確認できる．

表 3 感情強度推定の結果 (Closed Task)

Precision	J	Sa	Ant	Su	Ang	F	D	T
BL	0.43	0.43	0.40	**0.47**	0.33	0.38	0.39	0.36
M-SE	0.47	0.41	0.42	0.44	0.30	0.25	0.42	0.37
M-WID	0.49	0.46	**0.48**	0.38	0.39	**0.42**	**0.65**	**0.48**
M-SE-WID	**0.54**	**0.47**	0.46	**0.47**	**0.44**	0.21	0.46	**0.48**
Recall	J	Sa	Ant	Su	Ang	F	D	T
BL	0.45	0.43	0.40	0.37	0.29	0.33	0.34	0.35
M-SE	0.47	0.40	0.42	0.30	0.27	0.25	0.37	0.36
M-WID	0.48	0.44	0.47	**0.38**	0.32	**0.36**	**0.40**	**0.46**
M-SE-WID	**0.53**	**0.45**	**0.48**	0.35	**0.38**	0.21	**0.40**	**0.46**
F-measure	J	Sa	Ant	Su	Ang	F	D	T
BL	0.43	0.39	0.36	**0.44**	0.27	**0.32**	0.36	0.35
M-SE	0.45	0.39	0.37	0.28	0.27	0.24	0.37	0.36
M-WID	0.49	**0.46**	0.43	0.33	0.30	**0.32**	0.40	0.44
M-SE-WID	**0.51**	**0.46**	**0.45**	0.41	**0.33**	0.23	**0.42**	**0.45**
Accuracy	J	Sa	Ant	Su	Ang	F	D	T
BL	0.62	0.63	0.56	0.73	**0.87**	**0.82**	0.74	0.75
M-SE	0.63	0.62	0.57	0.72	0.86	**0.82**	0.78	0.76
M-WID	0.63	0.64	0.60	0.69	**0.87**	**0.82**	**0.80**	**0.80**
M-SE-WID	**0.69**	**0.65**	**0.63**	**0.74**	**0.87**	0.75	0.79	**0.80**

J=Joy, Sa=Sadness, Ant=Anticipation, Su=Surprise,
Ang=Anger, F=Fear, D=Disgust, T=Trust

Open Task における感情強度推定モデルの評価結果を表 4 に示す．まず，感情顕現性判定モデルの効果について検証する．F 値ならびに正解率について M-SE と BL，M-SE-WID と M-WID を比較すると，大部分の感情クラスにおいて，感情顕現性判定モデルを組み込んだモデル (M-SE, M-SE-WID) の方が評価値が高かった．再現率の結果を見ると，8 つのうち 6 つの感情クラスで M-SE が最高の成績を収めている．感情顕現性判定モデルの導入は特に再現率を向上させる効果があり，感情強度推定の性能の向上に貢献すると言える．

次に，書き手 ID の効果について検証する．F 値ならびに正解率について M-WID と BL，M-SE-WID と M-SE を比較すると，書き手 ID を使用するモデルが使用しないモデルを上回るのはおよそ半数の感情クラスに留まる．Closed Task の設定とは異なり，書き手 ID を感情強度推定モデルに組み込むことの明確な有効性は確認できなかった．これは，Open Task では，テストデータの書き手 ID は訓練データで最も類似した書き手の ID を代用しているため，書き手の情報をモデルに正確に組み込めていないためと考えられる．

Open Task と Closed Task の実験結果を比較すると，全般的に Open Task の方が評価値が低い．これは，書き手が未知である Open Task は Closed Task と比べて難しいことを裏付けるものである．

表 4 感情強度推定の結果 (Open Task)

Precision	J	Sa	Ant	Su	Ang	F	D	T
BL	0.42	**0.35**	0.38	0.39	**0.41**	0.45	**0.42**	0.30
M-SE	0.43	**0.35**	0.38	0.39	0.33	0.42	0.23	0.39
M-WID	**0.45**	0.30	0.34	0.38	0.26	0.41	0.37	0.27
M-SE-WID	0.39	0.34	**0.42**	**0.41**	0.33	**0.48**	0.38	**0.42**
Recall	J	Sa	Ant	Su	Ang	F	D	T
BL	0.39	0.34	0.35	0.30	0.30	0.31	**0.32**	0.28
M-SE	**0.41**	**0.39**	**0.36**	**0.32**	**0.36**	0.30	0.27	**0.34**
M-WID	**0.41**	0.30	0.35	0.28	0.31	0.29	0.26	0.28
M-SE-WID	0.39	0.34	0.35	0.31	0.35	**0.32**	0.28	**0.34**
F-measure	J	Sa	Ant	Su	Ang	F	D	T
BL	0.39	0.35	**0.35**	0.28	0.30	0.31	**0.29**	0.28
M-SE	**0.41**	**0.37**	**0.35**	**0.31**	**0.35**	0.31	0.25	**0.36**
M-WID	0.40	0.29	0.32	0.27	0.28	0.30	0.24	0.27
M-SE-WID	0.38	0.33	0.32	0.30	0.34	**0.33**	0.28	**0.36**
Accuracy	J	Sa	Ant	Su	Ang	F	D	T
BL	0.59	0.64	0.58	0.62	0.88	0.79	**0.79**	0.81
M-SE	0.62	**0.70**	0.59	0.64	**0.90**	0.84	0.65	0.69
M-WID	**0.63**	0.56	0.54	**0.68**	0.87	**0.86**	0.68	0.80
M-SE-WID	0.61	0.67	**0.63**	0.65	**0.90**	0.84	0.72	**0.90**

J=Joy, Sa=Sadness, Ant=Anticipation, Su=Surprise,
Ang=Anger, F=Fear, D=Disgust, T=Trust

5 おわりに

本論文では，書き手の感情強度を推定することを目的とし，書き手の感情がどれだけテキストの表層に明示的に出現しているかを示す情報として感情顕現性を，書き手の個性を示す情報として書き手 ID を，感情強度推定モデルに組み込む手法を提案した．今後の課題として，現在は感情顕現性を二値で表現しているが，書き手と読み手の感情強度の差を定量化し，感情顕現性の強さを学習することが挙げられる．また，Open Task では未知の書き手と類似した書き手を見つけるのが困難であったため，これを改善する手法や，書き手 ID の付与とは別のやり方で書き手の個人的な特徴をモデルに反映させる手法を探究することも重要な課題である．

参考文献

[1] Abdullah Alsaedi, Stuart Thomason, Floriana Grasso, and Phillip Brooker. Transfer learning model for social emotion prediction using writers emotions in comments. In **21st IEEE International Conference on Machine Learning and Applications (ICMLA 2022)**, pp. 396–400, 2022.

[2] Sven Buechel and Udo Hahn. EmoBank: Studying the impact of annotation perspective and representation format on dimensional emotion analysis. In **Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics**, pp. 578–585, 2017.

[3] Tomoyuki Kajiwara, Chenhui Chu, Noriko Takemura, Yuta Nakashima, and Hajime Nagahara. WRIME : A new dataset for emotional intensity estimation with subjective and objective annotations. In **Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies**, pp. 2095–2104, 2021.

[4] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In **Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies**, pp. 4171–4186, 2019.

[5] Hassan Alhuzali and Sophia Ananiadou. SpanEmo: Casting multi-label emotion classification as span-prediction. arXiv, 2101.10038, 2021.

[6] 鈴木陽也, 秋山和輝, 梶原智之, 二宮崇, 武村紀子, 中島悠太, 長原一. 書き手の性格情報を用いた感情強度推定. 人工知能学会全国大会論文集 (第 36 回), 2022. 4D3-GS-6-04.

[7] Lewis R. Goldberg. The development of markers for the Big-Five factor structure. **Psychological Assessment**, Vol. 4, No. 1, pp. 26–42, 1992.

[8] Robert Plutchik. A general psychoevolutionary theory of emotion. In Robert Plutchik and Henry Kellerman, editors, **Theories of Emotion**, pp. 3–33. Academic Press, 1980.

[9] BERT base Japanese (IPA dictionary, whole word masking enabled) – Hugging Face, (2024-1 閲覧). https://huggingface.co/cl-tohoku/bert-base-japanese-whole-word-masking/.