多ラベル分類モデルによる現代日本語短歌の感情識別

黒川真琳 [1]　持橋大地 [2,3]　小林一郎 [1]
[1] お茶の水女子大学 [2] 統計数理研究所 [3] 国立国語研究所
{g2120514,koba}@is.ocha.ac.jp daichi@ism.ac.jp

概要

本研究では，現代日本語短歌を対象とした感情識別タスクに対し，多ラベル分類モデルに基づくアプローチを検討する．現代日本語短歌および平文からなるデータセットを構築し，80 種類の感情ラベルによる客観感情アノテーションを付与した．これらのデータに対して Transformer ベースの分類モデルを適用し，感情識別を行った．実験では，短歌と平文の分類性能を比較し，短歌特有の表現が感情識別に与える影響を分析した．その結果，本手法は現代日本語短歌に対して一定の感情分類性能を示すことが確認された．

1　はじめに

感情分類は，自然言語処理における基盤的課題の一つであり，事前学習言語モデルの発展により，多ラベル感情分類においても高い性能が報告されている．しかし，これらの研究は主として日常的な文章を対象としており，短く文学的な表現を含むテキストに対する検討は十分とは言えない．

現代日本語短歌は，31 音という制約された形式の中で情景や心情を表現する短詩であり，感情が明示的な感情語として現れにくいという特徴を持つ．このため，短歌に対する感情識別は，従来の感情分類手法にとって依然として課題が残されている．

本研究では，現代日本語短歌を対象に，多ラベル分類モデルを用いた感情識別の適用可能性を検討する．短歌および平文からなるデータセットを用い，両者の分類性能を比較することで，短歌における感情識別の特性を明らかにすることを目的とする．

2　関連研究

感情分類は，テキストに内在する感情状態を推定する自然言語処理の基盤的課題であり，近年では大規模データセットと機械学習手法に基づく研究が主流となっている．特に，複数の感情が同時に表出し得ることを考慮した多ラベル感情分類は，GoEmotions データセットの提案により大規模に検証されている [1]．

また，Transformer に基づく事前学習言語モデルの発展により，感情分類性能は大きく向上した．BERT [2] や RoBERTa [3]，さらに多言語モデルである XLM-RoBERTa [4] は，日本語を含む感情分類タスクへの応用可能性を示している．しかし，これらの研究の多くは，比較的長く感情が明示的に表現されるテキストを対象としている．

一方，詩や文学作品を対象とした自然言語処理研究は，意味解析やスタイル分析を中心に行われてきた．Kao と Jurafsky は，英語詩の計算的分析を通じて，詩的スタイルと語彙・構文的特徴の関係を示している [5]．文学テキストにおける感情理解は，比喩や象徴表現に依存するため困難であり，Jacobs は，文学作品が読者に多層的な感情反応を引き起こすことを論じている [6]．そのため，日本語短歌を対象とした感情分類研究は依然として限られている．

以上より，感情分類研究は主として日常的テキストを対象として発展してきた一方で，短歌のような極短文かつ文学的表現を含むテキストに対する検討は十分ではない．本研究は，現代日本語短歌を対象に，多ラベル分類モデルと客観感情アノテーションを用いた感情識別を行い，感情分類研究と文学テキスト解析の接点を補完することを目的とする．

3　研究概要

本研究では，現代日本語短歌を対象とした感情分類タスクにおいて，多ラベル分類モデルに基づくアプローチを採用する．研究全体の流れを図 1 に示す．

まず，短歌および平文からなるテキストに対して感情アノテーションを付与し，感情分類データセットを構築する．次に，Transformer ベースの分類モデルを用いて学習を行い，短歌と平文の分類性能を比較することで，文種の違いが感情識別に与える影響

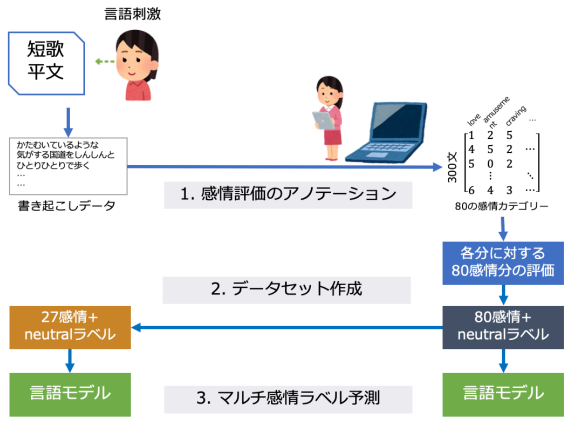

図 1　研究概要

を分析する．

本研究は，新たな分類手法の提案を目的とするものではなく，既存の感情分類モデルを短歌という言語資源に適用し，その有効性と課題を整理する点に主眼を置く．これにより，感情分類研究における対象テキストの拡張可能性を示すことを目指す．

4　データセットとアノテーション

4.1　感情収集対象となる短歌

本研究では，短歌が誘発する感情を分析するために，現代短歌と比較用の平文を収集した．短歌データは『現代日本語書き言葉均衡コーパス（BCCWJ）』，『桜前線開架宣言』，および『塔』から抽出した 150 首を用いた．また，比較対象として，BCCWJ に含まれる，短歌と同じく 31 文字程度の一般的な文（平文）150 文を選定した [7, 8]．

- 現代短歌の例：
  行ってきますと自分に伝えただいまと自分を迎える単身赴任
- 平文の例：
  水がたまった長い坑道を坑内員は休みながら進む

4.2　アンケートの実施

本研究では，短歌および平文からなる計 300 文を対象として，言語刺激に曝された際に誘発される感情についてアノテーションを収集した．評価対象とする感情カテゴリは，Koide-Majima ら [9] による日本語感情解析で用いられた 80 種類の感情カテゴリに基づく．各感情カテゴリの詳細は付録 表 5 に示す．

アノテーションは，日本語を母国語とするアノテータを対象に実施した．アノテータには，各文から読み取れる感情が，指定された感情カテゴリとどの程度一致しているかを，0（全く読み取れない）から 6（非常に強く読み取れる）までの 7 段階で評価するよう指示した．評価に際しては，アノテータ自身の主観的感情ではなく，テキスト表現から第三者として読み取れる感情に基づいて判断するよう求めた．

各感情カテゴリについて 4 人の異なるアノテータによる独立した評価を収集した．80 種類の感情カテゴリに対して 4 つの評価を得るため，クラウドソーシングサービス・ランサーズ[1)] を用いて，延べ 384 名（重複あり）のアノテータを募集した．1 回のアンケートでは，短歌・平文それぞれ 25 文を提示し，各文について 5 つの感情カテゴリを評価対象とした．この設計により，各文に対して多様な視点からの評価を収集した．

収集された感情アノテーションは，各文・各感情カテゴリごとに保存し，以降のデータセット構築に用いた．

4.3　データセット作成

本研究では，上述の感情強度アノテーションを，多ラベル感情分類タスクに適した形式へ変換する．

まず，各感情カテゴリについて，アノテータ間の評価のばらつきを標準偏差により算出する．標準偏差が所定の閾値を超える場合には，当該感情が文から安定して読み取られていないと判断し，その感情の強度を 0 とする．一方，標準偏差が閾値以下の場合には，アノテータの平均値を当該感情の強度として採用する．

次に，各文において平均強度が相対的に高い感情を選択し，上位の感情を最大 5 つまでラベルとして付与する．いずれの感情も付与条件を満たさない場合には，当該文に中立ラベルを付与する．これにより，各文は複数の感情ラベル，もしくは中立ラベルを持つものとして扱われる．

本研究では，感情カテゴリ数の異なる 2 種類のラベル設定を用いて実験を行う．1 つ目は，本研究で構築した 80 種類の感情カテゴリに中立ラベルを加えた「80 感情＋中立ラベル」の設定である．2 つ目は，既存研究との比較を目的として，GoEmotions データセットで定義されている 27 種類の感情カテ

1）　https://www.lancers.jp/

ゴリに中立ラベルを加えた「27 感情＋中立ラベル」の設定である．

80 感情から 27 感情への削減を行った理由は主に二つある．第一に，本研究で扱う短歌および平文のデータ数は 300 文と限定的であり，80 感情という細粒度なラベル設定では，各感情の出現頻度が極端に低くなるものが多く，学習の不安定化や性能評価のばらつきを招く可能性がある．第二に，GoEmotions は多ラベル感情分類において広く用いられている既存データセットであり，感情カテゴリを揃えることで，先行研究との比較可能性を確保することができる．

この設定では，80 種類の感情カテゴリを，感情の意味的近接性に基づく概念的な対応づけにより，GoEmotions の 27 感情カテゴリへ統合した．具体的には，各感情カテゴリを意味的に最も近いと判断される GoEmotions のカテゴリへ割り当て，対応が困難な感情については除外した．この対応づけは厳密な一対一対応を仮定するものではなく，概念レベルでの大まかなカテゴリ統合である．ラベル削減後も，前述の手順に従って，各文に対する多ラベル付与および中立ラベルの付与を行った．

5 実験

5.1 タスク詳細

本研究では，テキスト x に対して複数の感情が同時に付与され得る多ラベル感情分類タスクを扱う．感情集合を $\mathcal{E} = \{e_1, e_2, \ldots, e_K\}$ とし、本研究では $K = 28, 81$ とする．各テキストに対して，感情 e_k が読み取れるか否かを二値ラベル $y_k \in \{0, 1\}$ として予測する．モデルは入力テキスト x に対して，各感情ラベルごとの予測確率 $\hat{y}_k \in [0, 1]$ を出力し，これらを独立な二値分類問題の集合として扱う．

5.2 モデル

感情分類モデルには，Transformer に基づく事前学習済み言語モデル XLM-RoBERTa の“xlm-roberta-base”を使用する．入力テキストはトークナイズされた後，事前学習済みモデルに入力され，最終層の文表現を取得する．この文表現を線形変換層に入力し，各感情ラベルに対応する出力を得る．最終層では各感情ラベルに対してシグモイド関数を適用し，損失関数には多ラベル分類に適した二値交差エントロピー損失を用いる．

5.3 実験設定

本研究では，小規模データ環境における学習安定性と性能への影響を分析するため，以下の条件について比較実験を行った．

- ファインチューニングなし（事前学習モデルの直接適用）
- 80 感情＋中立ラベルによるファインチューニング
- 27 感情＋中立ラベルによるファインチューニング
- 更新層の違いによる比較
- 文表現（CLS / mean pooling）の違い
- 学習エポック数の違い

これらの条件を組み合わせ，文種別および設定別の性能変化を分析する．

5.4 評価指標

感情分類性能の評価には，micro-F1 スコアおよび macro-F1 スコアを用いる．micro-F1 は，すべてのラベルを通して真陽性・偽陽性・偽陰性を集計する指標であり，出現頻度の高い感情ラベルの影響を強く反映する．一方、macro-F1 は，各感情ラベルごとに算出した F1 スコアの平均であり，出現頻度の低い感情ラベルを含めた全体的な性能を評価する指標である．本研究では，感情ラベル間の出現頻度に偏りが存在することから，両指標を併用することで，多ラベル感情分類性能を多角的に評価する．なお，本研究では交差検証は行わず，事前に分割した学習データおよび評価データを用いて性能評価を行う．

5.5 結果

表 1 に，ラベル設定およびファインチューニング条件の違いによる性能を示す．ファインチューニングを行わない場合，性能は著しく低く，事前学習モデルのみでは本タスクへの適応が困難であることが分かる．80 感情＋中立ラベルによるファインチューニングでは性能が向上し，さらに 27 感情＋中立ラベルに削減することで，micro-F1，macro-F1 ともに最も高い値を示した．

以降は 27 感情＋中立ラベルによるファインチューニングの比較を行う．

表 2 に，更新層の違いによる性能比較を示す．最終層のみを更新した場合と比較して，上位数層

表 1　ラベル設定およびファインチューニング条件の違いによる性能比較

設定	micro-F1	macro-F1
w/o fine-tuning	0.092	0.062
80 Labels + “neutral”	0.146	0.229
27 Labels + “neutral”	0.271	0.318

または全層を更新した設定では，全体として性能の向上が確認された．特に，上位数層を更新した設定では，短歌・平文の双方を含む条件において，macro-F1 が最も高い値を示した．

表 2　更新層の違いによる感情分類性能

更新層設定	micro-F1	macro-F1
最終層のみ（短歌）	0.264	0.286
最終層のみ（平文）	0.191	0.163
最終層のみ	0.231	0.253
上位数層（短歌）	0.288	0.286
上位数層（平文）	0.249	0.275
上位数層	0.271	0.317
全層（短歌）	0.292	0.257
全層（平文）	0.242	0.206
全層	0.270	0.313

表 3 に，文表現の集約方法の違いによる性能を示す．CLS トークンおよび Mean pooling のいずれを用いた場合も，同程度の性能が得られたが，CLS トークンは平文，Mean pooling は短歌においてやや高い性能を示した．

表 3　文表現の集約方法による感情分類性能

文表現	micro-F1	macro-F1
CLS （短歌）	0.295	0.322
CLS（平文）	0.276	0.313
CLS	0.286	0.338
Mean pooling （短歌）	0.299	0.336
Mean pooling（平文）	0.263	0.304
Mean pooling	0.284	0.340

表 4 は，学習エポック数の違いによる性能を示す．20 エポックで最も高い性能が得られ，それ以上学習を行うと性能が低下する傾向が見られた．

5.6　考察

本研究の結果から，現代日本語短歌を対象とした感情識別においては，ラベル設計および学習条件の選択がモデル性能に大きく影響することが示された．特に，小規模データ環境では，感情カテゴリを過度に細分化しない設定が，学習の安定性という観点から重要であると考えられる．

表 4　エポック数の違いによる性能比較

エポック数	micro-F1	macro-F1
10	0.271	0.317
20	0.286	0.338
30	0.274	0.324

また，事前学習モデルの更新範囲を適切に制御することで，短歌に含まれる文学的・暗示的表現に対しても，表現の再調整が一定程度有効に働くことが示唆された．一方で，更新範囲を広げすぎると安定性を損なう可能性もあり，小規模データでは段階的な更新が妥当であると考えられる．

文表現の集約方法や学習エポック数については，いずれも性能に影響を与える要因であるものの，極端な設定は過学習を招きやすいことが確認された．以上より，短歌対を対象とした感情識別において，モデル構成そのものよりも，学習条件の設計が性能を左右する重要な要素であるといえる．

6　まとめ

本研究では，現代日本語短歌を対象とした多ラベル感情識別タスクに対し，Transformer ベースの分類モデルを適用し，その有効性と課題を体系的に検討した．短歌および平文からなるデータセットを構築し，客観感情アノテーションに基づく多ラベル感情分類実験を行った点に本研究の特徴がある．

実験の結果，感情ラベルの粒度や学習条件の設計が分類性能に大きく影響することが示され，短歌のように感情が明示されにくいテキストに対しても，適切な設定の下では多ラベル感情分類が一定程度可能であることが確認された．特に，小規模データ環境においては，感情カテゴリの統合や更新層の制御が，学習の安定性と性能向上の両立に寄与することが示唆された．

本研究は，感情分類研究の主な対象であった日常的テキストとは異なる言語資源に対して，既存の分類モデルを適用し得る可能性とその限界を整理した点で意義を有する．今後は，比喩や情景描写といった短歌特有の表現を考慮した表現学習の導入や，感情強度を直接予測する手法との比較を通じて，文学的テキストにおける感情表現理解の深化が期待される．

謝辞

本研究は ROIS-DS-JOINT2025 の支援を受けました．ここに深謝いたします．

参考文献

[1] Dorottya Demszky, et al. Goemotions: A dataset of fine-grained emotions. In **ACL**, 2020.
[2] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. In **NAACL**, 2019.
[3] Yinhan Liu, et al. Roberta: A robustly optimized bert pretraining approach. In **arXiv preprint arXiv:1907.11692**, 2019.
[4] Alexis Conneau, et al. Unsupervised cross-lingual representation learning at scale. In **ACL**, 2020.
[5] Justine T. Kao and Dan Jurafsky. A computational analysis of poetic style. In **NAACL**, 2015.
[6] Arthur M. Jacobs. Neurocognitive poetics: Methods and models for investigating the neuronal and cognitive-affective bases of literature reception. **Frontiers in Human Neuroscience**, 2018.
[7] 一般社団法人塔短歌会. 塔. Vol. 63, No. 4, 2016.
[8] 山田航. 桜前線開架宣言. 左右社, 2015.
[9] N. Koide-Majima, T. Nakai, and S. Nishimoto. Distinct dimensions of emotion in the human brain and their representation on the cortical surface. **NeuroImage**, Vol. 222, p. 117258, 2020.

A 付録

表 5 80 の感情カテゴリ

love (愛)	joy (喜び)	calmness (穏やかさ)
sadness (悲しみ)	awe (畏怖)	interest (興味)
sexual desire (性的欲求)	tension (緊迫感)	awkwardness (気まずさ)
fear (恐怖)	happiness (幸福)	affection (愛情・好意)
emotional hurt (心の痛み)	empathy (共感)	unrest (動揺)
fever (熱狂)	positive-expectation (前向きな期待)	indecency (下品)
contempt (侮蔑)	positive-emotion (前向きな感情)	tenderness (優しさ)
relaxedness (リラックス)	negative-emotion (否定的感情)	protectiveness (保護)
cuteness (可愛らしさ)	annoyance (迷惑)	distress （苦痛）
amusement (娯楽)	nostalgia (懐かしさ)	relief (安堵)
admiration (賞賛)	confusion (混乱)	satisfaction (満足)
surprise (驚き)	anger (怒り)	disgust (嫌悪)
horror (bloodcurdling) (恐ろしさ)	friendliness (親しみやすさ)	aggressiveness (攻撃性)
liking (好感)	sympathy (同情)	compassion (思いやり)
exuberance (過剰な喜び)	scare (feel a cill) (ゾッとする)	throb (どきどき)
embarrassment (恥ずかしさ)	alertness (警戒心)	vigor (活気)
pensiveness (物思い)	acceptance (容認)	hostility (敵意)
elation (意気揚々)	attachment (愛着)	positive-fear (肯定的恐怖)
stress (ストレス)	craving (渇望)	boredom (退屈)
romance (ロマンス)	aesthetic appreciation (美的満足)	entrancement (魅了、没入)
excitement (興奮)	nervousness (神経質)	anxiety (不安)
empathic pain (共感性苦痛)	laughing (笑い)	ridiculousness (滑稽)
shedding tears (涙を流す)	lethargy (無気力)	curiousness (好奇心)
appreciation of beauty (美の鑑賞)	daze (ぼんやり)	sexiness (セクシーさ)
oddness (奇妙さ)	eeriness (不気味さ)	longing (憧れ)
melancholy (憂鬱)	unease (不安)	levity (平静)
coolness (冷静さ)	encouragement (励まし)	

表 6 27 の感情カテゴリ: GoEmotions

admiration (賞賛)	amusement (娯楽)	anger (怒り)
annoyance (迷惑)	approval (賛同)	caring (思いやり)
confusion (混乱)	curiosity (好奇心)	desire (欲求)
disappointment (失望)	disapproval (不承認・否定)	disgust (嫌悪)
embarrassment (恥ずかしさ)	excitement (興奮)	fear (恐怖)
gratitude (感謝)	grief (悲嘆)	joy (喜び)
love (愛)	nervousness (緊張・不安)	optimism (楽観)
pride (誇り)	realization (気づき・理解)	relief (安堵)
remorse (後悔)	sadness (悲しみ)	surprise (驚き)