

マルチタスク学習を用いた ドメインが異なるテキストにおける感情分析

浅野 晴暉[†] 松原 雅文[‡]

岩手県立大学大学院ソフトウェア情報学研究科[†] 岩手県立大学ソフトウェア情報学部[‡]

1. はじめに

テキスト感情分析はテキストから感情を機械的に識別する技術である。この技術には、喜・怒・哀・楽といった多種の感情を分析する手法が存在するが、現状ではネガティブ・ポジティブの2値を対象とした分析と比較すると感情の粒度が細かいため、高精度な分析が困難となっている。日本語において、従来から行われている多種の感情分析の多くは、感情語辞書を用いたルールベースの手法である。しかし、ルールベースの手法では文法規則や多義語の影響により、文脈の識別が難しいといった問題が存在する。テキストには様々な感情が含まれているため、感情極性の分析だけでは不十分なケースが存在する¹⁾。そのため、多種の感情分析精度を向上させることには意義があり、その実現には多量のデータセットや深層学習の技術が必要であると考えられる。しかし、感情は機械学習の分類対象の中でも比較的曖昧な概念である。そのため、感情分析は解釈の曖昧性から、データセットの作成に、よりコストがかかるタスクであると考えられる。

上記の背景から、絵文字を用いた Distant supervision により、感情分析における教師データを自動構築する研究が存在する。Distant supervision は知識ベースから得られる手掛かりをもとに生データから関係を抽出し、教師データを自動生成する手法である²⁾。絵文字が感情表現に用いられる性質を利用し、あらかじめ絵文字が表すと考えられる感情を定義する。そして、ある絵文字を含んだテキストは定義した感情を表すという仮定を置き、そのテキストに感情ラベルを付与するという手法である³⁾。この手法では、低コストで多量の教師データを自動構築できる。しかし、データ内にノイズが含まれることや、絵文字と感情との関係性の解釈に主観的な要素や曖昧性を含んでしまうことが問題となる。

そこで、我々は絵文字を用いて自動生成した教師データを適切に学習し、多種の感情分析精度を向上させることを目的とした手法を提案している⁴⁾。本手法では、絵文字が表すと考えられる感情を客観的な指標を用いて定義することで、良質な教師データを生成する。そして、そのデータを適切に学習するためにマルチタスク学習を行う。マルチタスク学習とは、複数のタスク

を共同で学習する手法である。この学習法は関連するタスクを共同で学習した場合に、特定、あるいは複数タスクの精度を向上させることが知られている。

テキストから絵文字が表すと考えられる感情を予測するタスクを絵文字予測タスクと定義し、これを感情分析タスクの補助タスクに用いる。この処理によって、定義した感情の主観性や曖昧性を含み重みに影響されることなく、感情分析に有益な情報のみを伝達させ、感情分析の性能を向上させることを目指している。

本稿では、絵文字から生成した教師データのドメインが解析対象と異なる場合における、本手法の有効性を検証する。絵文字を用いて自動生成した教師データと異なるドメインのデータを解析対象とし、モデルの評価を行うことで有効性を確認する。

2. 関連研究

2.1. 感情分析

日本語において、多種の感情を対象とした感情分析では感情語辞書が頻繁に用いられる。主に、「感情表現辞典」によるパターンマッチングやそれを拡張した手法である ML-Ask が提案されている^{5, 6, 7)}。一方で、英語など他言語を対象とした多種の感情分析の研究では、Transformer をはじめとした深層学習を用いた研究が盛んである^{8, 1)}。英語を対象とした感情の分析では、ブルチックの8感情などを対象にしている⁹⁾。

感情分析におけるリソース不足に言及し、絵文字を用いた Distant supervision により、ラベル付けのコストを削減しようとする研究が行われてきた。基本的には、絵文字が表す感情を定義し、絵文字を含むテキストに定義した感情ラベルを付与する手法である^{3, 10)}。感情を定義せずに絵文字を予測する過程で得られた重みを転移学習する研究も存在する¹¹⁾。

2.2. マルチタスク学習

昨今における深層学習の普及によりマルチタスク学習が盛んに活用されており、テキスト分類や感情分析においても適用事例が多数存在する¹²⁾。マルチタスク学習を行うモデルには、共有層とは別にタスク固有の層を設けることでタスク間で不変な空間と固有の空間に分離するモデルが存在する¹³⁾。このモデルは、共有層のみ用いたモデルと比較し、ドメインが異なるデータを適切に学習することができる。上記モデルに敵対的損失を加えることで、より正確に空間を分離することのできる、敵対的マルチタスク学習 (Adversarial Multi-Task Learning) という手法も存在する¹⁴⁾。敵対的マルチタスク学習は、近年では、アスペクトベースの感情分析

Sentiment Analysis in Texts with Different Domains using Multi-task Learning

Haruki ASANO[†], Masafumi MATSUHARA[‡]

[†]Graduate School of Software and Information Science, Iwate Prefectural University, [‡]Faculty of Software and Information Science, Iwate Prefectural University

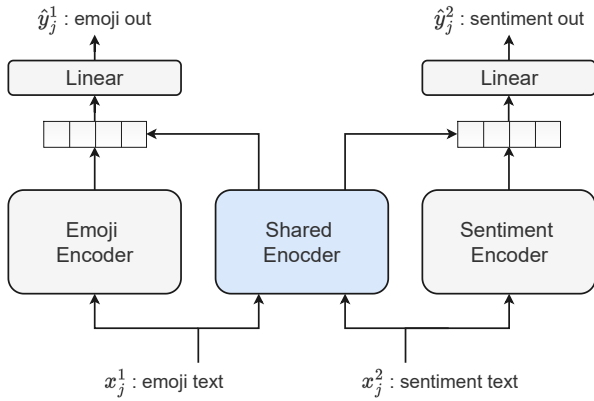


図 1: マルチタスク学習モデルのアーキテクチャに
応用されている¹⁵⁾.

3. 提案手法

3.1. 絵文字付きデータの収集・教師データの生成

絵文字は Unicode 文字として定義されているため、絵文字をクエリとしてパターンマッチングすることができる。そのため、絵文字付きのテキストを多数収集することは、容易な処理である。多量のテキストリソースから、絵文字付きテキストを抽出する手段として Twitter API¹⁾ が利用可能である。本研究では Twitter API を用いて対象となる絵文字を含むツイートを収集し、絵文字予測の学習に用いる教師データを生成する。

本手法では、Emotag1200 と呼ばれるデータセットをもとに絵文字の感情を定義する¹⁶⁾。Emotag1200 は 150 個の絵文字と 8 種類の感情 (喜び, 悲しみ, 期待, 驚き, 怒り, 恐れ, 嫌悪, 信頼) との連想性が、9 人の評価者によってアノテーションされている。Emotag1200 に基づき、絵文字を 8 種類の感情値に変換する。

3.2. マルチタスク学習の設定

本手法では絵文字予測と感情値予測の 2 タスクを対象とした、深層学習ベースのマルチタスク学習を行う。モデルのアーキテクチャを図 1 に示す。各タスクはサンプル数 n_t のトレーニングデータからなるデータセット \mathcal{D}_t を持つ。 t はタスクの id であり、絵文字予測を 1、感情値予測を 2 とする。すなわち、 $\mathcal{D}_t = \{x_j^t, y_j^t\}_{j=1}^{n_t}$ であり、 x_j^t はタスク t の j 番目の入力テキストサンプル、 y_j^t はそのラベルである。テキストの入力形状は $x_j^t \in \mathbb{R}^l$ であり、 l はシーケンス長、ベクトルの各要素は単語の index である。なお、シーケンスは任意の定数によって padding され、固定長となる。

テキストはトークナイズされた後、Shared Encoder とタスク固有の Encoder に入力され、それぞれの特徴量が算出される。その後、二つの特徴量を加算し全結合することで各タスクの出力を求める。感情分析は同一の単語、文章表現であってもドメインの違いにより意味合いが大きく異なるケースがある⁸⁾。そのため、仮に感情値予測と異なるドメインから収集した絵文字付きデータを用いる際、性能が低下してしまう恐れがあ

る。そこで、本手法では負の知識伝達による精度低下を防ぐために、共有層だけではなくタスク固有の層を追加している¹⁴⁾。

3.3. BERT による Text Encoding

本手法では Encoder に BERT (Bidirectional Encoder Representations from Transformers) を用いる¹⁷⁾。はじめに、入力 x_j^t は埋め込み次元 d_e に Word Embedding され $H \in \mathbb{R}^{l \times d_e}$ に変換される。位置エンコーディングと文種の判別を行った後、式 (1) により query, key, value ベクトルを算出する。

$$\begin{aligned} Q_i &= HW_i^Q \\ K_i &= HW_i^K \\ V_i &= HW_i^V \end{aligned} \quad (1)$$

i は Attention head のインデックスを表し、 $W_i^Q \in \mathbb{R}^{d_e \times d_q}$, $W_i^K \in \mathbb{R}^{d_e \times d_k}$, $W_i^V \in \mathbb{R}^{d_e \times d_v}$ は特徴量の次元を変換する重み行列である。そして、式 (2) により、Multi-Head Attention を計算する。

$$\begin{aligned} \text{Attention}(Q_i, K_i, V_i) &= \text{softmax}\left(\frac{Q_i^t K_i}{\sqrt{d_k}}\right) V_i \\ \text{MultiHead}(Q, K, V) &= \text{Concat}(\text{head}_1, \dots, \text{head}_h) W^O \\ \text{where head}_i &= \text{Attention}(Q_i, K_i, V_i) \end{aligned} \quad (2)$$

複数の head を Concat し、 $W^O \in \mathbb{R}^{hd_v \times d_e}$ により、もとの埋め込み次元に戻す。複数の正則化を適用し、最終的に GELU 関数にかけることで出力を得る。

なお、記述の簡略化のために、上記計算を層数分繰り返す BERT によるエンコードを式 (3) と定義する。

$$H = \text{BERT}(x_j^t) \quad (3)$$

3.4. 絵文字予測の学習

絵文字は多くの場合において感情を表す性質をもつ。そのため、テキストを入力とし、付与されている絵文字を予測するタスクからは、テキストと感情との関係性を表す知識表現が得られると考えられる¹¹⁾。

絵文字予測のアーキテクチャを図 2 に示す。Encoder に BERT を用いているため、入力するテキストの文頭は [CLS]、文末は [SEP] トークンとなる。予測するラベルは付与されていた絵文字の Emotag1200 の値である。そのため、絵文字予測と定義しているが、実際には絵文字から生成される感情値を予測するタスクである。単に絵文字を予測するのではなく、絵文字が表すと考えられる感情を予測することで絵文字と感情とのマッピングを明確にする狙いがある。

テキスト x_j^1 を絵文字予測固有の Encoder と感情値予測タスクと共有された Encoder にそれぞれ入力し、隠れ状態 $H_{\text{emoji}} \in \mathbb{R}^{l \times d_e}$ と $H_{\text{shared}} \in \mathbb{R}^{l \times d_e}$ を算出

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

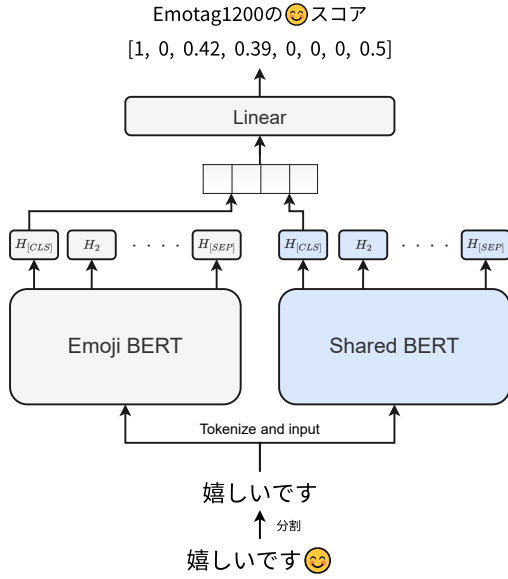


図 2: 絵文字予測の学習

する。

$$\mathbf{H}_{\text{emoji}} = \text{BERT}_{\text{emoji}}(x_j^1) \quad (4)$$

$$\mathbf{H}_{\text{shared}} = \text{BERT}_{\text{shared}}(x_j^1) \quad (5)$$

その後、それぞれの [CLS] に対応する隠れ状態を加算しテキスト特徴量 $C_{\text{emoji}} \in \mathbb{R}^{2d_e}$ を算出し、全結合することで出力を求める。

$$C_{\text{emoji}} = \text{Concat}(\mathbf{H}_{\text{emoji},[\text{CLS}]}, \mathbf{H}_{\text{shared},[\text{CLS}]}) \quad (6)$$

$$\hat{y}_j^1 = C_{\text{emoji}}W^1 + b^1 \quad (7)$$

3.5. 損失の算出

マルチタスク学習では、各タスクの損失が算出されるため、最終的な損失 L は、これらを合算することで求められる。本手法では式 (8) で最終損失を求める。

$$L = \lambda_1 L_e + \lambda_2 L_s \quad (8)$$

λ_t の重みパラメータによって絵文字予測の損失 L_e と、感情値予測の損失 L_s を調整する。

このように本手法においては、上記損失を共同で最適化することで、絵文字予測タスクで得られた知識表現を感情値予測に伝達する。得られた知識が感情分析に有益に作用し、感情分析の性能を向上させることが期待される。

4. 実験

4.1. データセット

感情分析のデータセットには WRIME¹⁸⁾ と Japanese Realistic Textual Entailment Corpus¹⁹⁾ を用いる。

WRIME は主観と客観両方の視点から、テキストに喜び、悲しみ、期待、驚き、怒り、恐れ、嫌悪、信頼の 8 感情が 4 段階 (0: 無・1: 弱・2: 中・3: 強) でラベル付けされている。客観の視点では、3 人の評価者にアノテーションされている。今回の実験には客観の視点で付与

表 1: 選定した絵文字とサンプル数

感情	絵文字	Name	総数
喜び	😊	smiling face	531
悲しみ	😭	crying face	2,389
驚き	!	exclamation mark	2,517
怒り	😡	pouting face	10,252
恐れ	😱	fearful face	4,034
嫌悪	👎	thumbs down	5,361
信頼	😊	kissing face with smiling eyes	9,799

された 3 人のラベル値を平均したものをを用い、データは最小値 0、最大値 1 に正規化する。また、学習セット、検証セット、テストセットの数はそれぞれ 30,000, 2,500, 2,500 となっている。

Japanese Realistic Textual Entailment Corpus はじゃらんに掲載されているホテルレビューから生成された感情分析データセットである。感情の種類はポジティブ、ネガティブ、ニュートラルとなっている。また、学習セット、検証セット、テストセットの数はそれぞれ 3,888, 1,112, 533 となっている。本研究では、複数の感情に焦点を当てているが、ドメインの異なるデータにおける有効性を示すために本データセットも評価対象として用いる。

絵文字予測のデータセットは対象となる絵文字を含むツイートを抽出することで生成される。今回は Emotag1200 の各感情カテゴリにおいて最も値が大きい絵文字を、その感情の代表絵文字とし対象にした。また、サンプルサイズに関しては、感情カテゴリが不均衡になることを防ぐためオーバーサンプリングを行った。データセット内で値が 1 以上のラベルを確認したところ、期待のラベルが最も多く 11,391 件であった。そこで、期待を除く各感情の代表絵文字を含むテキストを 11,391 とラベル数の差分だけ抽出した。対象とした絵文字、およびサンプル数を表 1 に示す。なお、ラベルの強弱は無視しており、絵文字には代表となる感情以外のスコアも付与されているため、各感情のラベル数や値は完全に均等になってはいない。

絵文字予測のデータセットと WRIME はどちらも SNS のテキストから作られているため、同一ドメインのデータセットとして捉えることができる。一方で、Japanese Realistic Textual Entailment Corpus ホテルのレビューテキストであるため、比較的フォーマルなテキストであると考えられる。そのため、絵文字予測のデータセットとは異なるドメインのデータセットとして捉えることができる。

4.2. 比較モデル

複数の感情を対象とする実験では、以下のモデルの性能を比較する。

- BERT_{BASE}: WRIME を用いて感情値予測の学習
- BERT_{AUG}: WRIME と絵文字予測データセットを用いて感情値予測の学習

表 2: 各モデルの P@k, R@k, F@k

model	P@1	P@2	P@3	R@1	R@2	R@3	F@1	F@2	F@3
BERT _{BASE}	.698	.572	.467	.443	.672	.790	.542	.618	.587
BERT _{AUG}	.678	.535	.446	.427	.627	.756	.524	.577	.561
BERT _{SP-MTL}	.724	.581	.471	.458	.680	.799	.561	.626	.593

- BERT_{SP-MTL}: WRIME と絵文字予測データセットを用いて、共有層とタスク固有の層でマルチタスク学習 (提案手法)

BERT_{AUG} は WRIME と絵文字予測データセットを結合したものを感情値予測として学習したモデルである。絵文字予測データセットは Emotag1200 によって、WRIME と同カテゴリの感情値に変換されているため、結合することができる。このモデルは、マルチタスク学習のモデルと異なり、絵文字予測データセットのラベルが WRIME と同等に学習される。WRIME は $[0, 3]$ のラベル y が $[0, 1]$ に正規化されており、Emotag1200 も同様に $[0, 1]$ である。そのため、Emotag1200 のラベル y' と y の対応関係は $y = 3y'$ になっている。

ネガティブ・ポジティブを対象とする実験では、以下のモデルの性能を比較する。

- BERT_{BASE}: Japanese Realistic Textual Entailment Corpus を用いて感情値予測の学習
- BERT_{SP-MTL}: Japanese Realistic Textual Entailment Corpus と絵文字予測データセットを用いて、共有層とタスク固有の層でマルチタスク学習 (提案手法)

なお、いずれのモデルにおいても東北大学の乾研究室で公開されている日本語 BERT 訓練済みモデル²を用いる。

4.3. 学習設定

Encoder に用いる BERT は、層数: 12, 隠れ状態の次元: 768, Attention head の数: 12 となっており、シーケンス長は 64 としている。オプティマイザーには AdaBound²⁰ を用い、学習率の初期値は 1×10^{-5} , 最終学習率は 1×10^{-2} である。BERT_{SP-MTL} の学習損失に用いる重みパラメータ λ_1, λ_2 は、学習の収束が均一になるように調整し、それぞれ 0.4, 0.6 に設定している。絵文字予測の出力には恒等関数を用い、損失関数には MSE (Mean squared error) を用いる。WRIME の学習時の出力にはシグモイド関数, Japanese Realistic Textual Entailment Corpus の学習時の出力には Softmax 関数を用いる。損失関数にはいずれも、Binary cross entropy を用いる。

4.4. 評価指標

WRIME に対する評価には P@k, R@k, F@k を用いる。マルチラベル分類においては、ラベルの次元数

表 3: Precision, Recall, F 値のマクロ平均

model	Precision	Recall	F 値
BERT _{BASE}	.85	.79	.82
BERT _{SP-MTL}	.85	.82	.83

とは対照的に、各インスタンスには少量のラベルしか付与されていない場合が多い。そのため、上位何件かを対象にしたランク評価が用いられる²¹。WRIME のデータにおいても、重要な感情は上位 1~3 件ほどであると考えられる。そのため、重要な感情の識別力を把握するためにランク指標を用いる。

Japanese Realistic Textual Entailment Corpus に対する評価には、ポジティブ, ニュートラル, ネガティブにおける、Precision, Recall, F 値のマクロ平均を用いる。

4.5. 結果および考察

表 2 に各モデルの P@k, R@k, F@k を示す。表 2 より、BERT_{SP-MTL} のランク指標が、BERT_{BASE}, BERT_{AUG} と比較し、すべて向上していることが確認できる。上記の結果より、重要な感情を識別する能力がマルチタスク学習によって向上したことがわかる。一方で、BERT_{AUG} は他のモデルと比較し、性能が著しく低下している。この結果は、WRIME と絵文字予測データセット間におけるラベルの対応関係が適切でなかったことが原因であると考えられる。また、絵文字予測データ内に含まれるノイズに過剰に適合してしまったことも原因であると考えられる。ラベルの重みをチューニングすることで性能が改善される可能性もあるが、ノイズのフィルタリングやスケールリングを含め、感情分析に有益な知識だけを伝達することのできるマルチタスク学習の方が、有効な学習法であると考えられる。

表 3 に各モデルの Precision, Recall, F 値のマクロ平均を示す。表 3 より、BERT_{SP-MTL} における性能の向上が確認できる。このように、絵文字予測データのドメインが解析対象と異なる場合においても本手法が有効であることが確認された。

今回の実験では、ネガティブ・ポジティブを解析対象とする場合でも、WRIME の学習と同様の絵文字予測データセットを用いている。絵文字は 8 種の感情に変換されているが、この 8 種の感情を予測するタスクからネガティブ・ポジティブ分類に有益な知識表現を得ることができた。よって、今後サンプル数やラベルの表現をネガティブ・ポジティブのデータに最適化することでさらに精度が向上する可能性がある。

²<https://huggingface.co/cl-tohoku/bert-base-japanese-whole-word-masking>

5. おわりに

感情分析においては、解釈の曖昧性からデータセットの作成にコストがかかる。そこで、我々は絵文字を用いて教師データを自動生成し、それをマルチタスク学習に適用させる手法を提案している。具体的には、タスク固有のBERTとタスク間で共有されたBERTを設け、それぞれのEncoderから得られる特徴量を結合して、感情値と絵文字を予測するモデルを構築した。本稿では、モデルの比較実験の結果より絵文字から生成した教師データのドメインが解析対象と異なる場合においても本手法が有効であることを示した。

今後は、絵文字のサンプル数や変換する感情ラベルを変更し、よりモデルの性能を向上させることのできる絵文字の種類やサンプル数、および感情ラベルを検討していく予定である。

謝辞

本研究の一部は JSPS 科研費 21K12611 の助成を受けたものである。

参考文献

- 1) Nafis Irtiza Tripto and Mohammed Eunus Ali. Detecting multilabel sentiment and emotions from bangla youtube comments. In *2018 International Conference on Bangla Speech and Language Processing (ICBSLP)*, pp. 1–6, 2018.
- 2) Mike Mintz, Steven Bills, Rion Snow, and Daniel Jurafsky. Distant supervision for relation extraction without labeled data. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, pp. 1003–1011, Suntec, Singapore, August 2009.
- 3) 林友超, 崔鴻翌, 宇津呂武仁. ツイート中の主観極性同定モデルの訓練事例収集における絵文字の利用. 知能と情報, Vol. 32, No. 5, pp. 923–933, 2020.
- 4) 浅野晴暉, 松原雅文. 絵文字予測を感情分析の補助タスクとしたマルチタスク学習手法. 第9回自然言語処理シンポジウム, Vol. 2022-NL-254, No. 2, Nov 2022.
- 5) Michal Ptaszynski, Pawel Dybala, Rafal Rzepka, Kenji Araki, and Fumito Masui. ML-ask: Open source affect analysis software for textual input in japanese. *Journal of Open Research Software*, Vol. 5, No. 1, p. 16, 2017.
- 6) 鳥海不二夫, 榊剛史, 吉田光男. ソーシャルメディアを用いた新型コロナ禍における感情変化の分析. 人工知能学会論文誌, Vol. 35, No. 4, pp. F-K45–1–7, 2020.
- 7) 藤田俊, 鷹野孝典. コンテンツのユーザに与える感情を考慮したコンテンツ推薦手法の検討. 情報処理学会論文誌, Vol. 61, No. 6, pp. 1200–1209, jun. 2020.
- 8) Wenhao Ying, Rong Xiang, and Qin Lu. Improving multi-label emotion classification by integrating both general and domain-specific knowledge. In *Proceedings of the 5th Workshop on Noisy User-generated Text (W-NUT 2019)*, pp. 316–321, Hong Kong, China, November 2019.
- 9) Neel Kant, Raul Puri, Nikolai Yakovenko, and Bryan Catanzaro. Practical text classification with large pre-trained language models, 2018.
- 10) Jared Suttles and Nancy Ide. Distant supervision for emotion classification with discrete binary values. In *International Conference on Intelligent Text Processing and Computational Linguistics*, pp. 121–136. Springer, 2013.
- 11) Bjarke Felbo, Alan Mislove, Anders Søgaard, Iyad Rahwan, and Sune Lehmann. Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pp. 1615–1625, Copenhagen, Denmark, September 2017.
- 12) Jian Zhang, Ke Yan, and Yuchang Mo. Multi-task learning for sentiment analysis with hard-sharing and task recognition mechanisms. *Information*, Vol. 12, No. 5, 2021.
- 13) Pengfei Liu, Xipeng Qiu, and Xuanjing Huang. Recurrent neural network for text classification with multi-task learning. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, IJCAI'16*, pp. 2873–2879. AAAI Press, 2016.
- 14) Pengfei Liu, Xipeng Qiu, and Xuanjing Huang. Adversarial multi-task learning for text classification. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1–10, Vancouver, Canada, July 2017.
- 15) Bin Liang, Rongdi Yin, Lin Gui, Jiachen Du, Yulan He, and Ruifeng Xu. Aspect-invariant sentiment features learning: Adversarial multi-task learning for aspect-based sentiment analysis. New York, NY, USA, 2020. Association for Computing Machinery.
- 16) Abu Awal Md Shoeb and Gerard de Melo. Emo-Tag1200: Understanding the association between emojis and emotions. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 8957–8967, Online, November 2020.
- 17) Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pp. 4171–4186, Minneapolis, Minnesota, June 2019.
- 18) Tomoyuki Kajiwar, 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, Online, June 2021.
- 19) Yuta Hayashibe. Japanese realistic textual entailment corpus. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pp. 6827–6834, Marseille, France, May 2020. European Language Resources Association.
- 20) Liangchen Luo, Yuanhao Xiong, Yan Liu, and Xu Sun. Adaptive gradient methods with dynamic bound of learning rate. In *Proceedings of the 7th International Conference on Learning Representations*, New Orleans, Louisiana, May 2019.
- 21) Jingzhou Liu, Wei-Cheng Chang, Yuexin Wu, and Yiming Yang. Deep learning for extreme multi-label text classification. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '17*, New York, NY, USA, 2017.