

Tweet 日本語感情コーパスに対する BERT の効果の検証

池上 達也[†] 任 福継[‡] 西出 俊[‡] 康 鑫[‡]徳島大学創成科学研究科[†] 徳島大学大学院社会産業理工学研究部[‡]

1. はじめに

Twitter では、多くのユーザが多様な言語表現でコミュニケーションをとっている。そのため、Twitter 上の宣伝情報を対象としたユーザの議論や感情への影響の分析[1]など、Twitter を利用した感情分析研究は多くの成果を示している。

また、双方向トランスフォーマーエンコーダとして登場した BERT[2]は、テキストを対象とした多くの感情分類研究で使用され、従来の手法よりも優れた結果を示している。

よって、本研究ではその流れを踏襲し、Tweet 日本語感情コーパスに対して、BERT におけるモデルの構成・学習法と感情分類の関係性の解析を目的とする。

2. モデルの構造

モデルの構造を図 1に示す。本モデルは、8 つの感情(不安, 怒り, 悲しみ, 嫌悪, 喜び, 愛, 期待, 驚き)と無感情の合計 9 つの感情を分類することを目的とする。活性化関数に Tanh 関数, 損失関数には, BCEWithLogitsLoss 関数, オプティマイザには, AdamW をそれぞれ使用した。

本研究では、事前学習済み BERT モデルとして東北大学で公開されており、日本語版 Wikipedia で訓練されているモデルの「cl-tohoku/bert-base-japanese」を使用する。形態素解析には MeCab を使用し、形態素解析された単語ごとにトークン化し得られた単語の埋め込み表現を BERT に入力することで、1 つのテキストの特徴として得られる 768 次元のベクトルを利用し、感情分類を行う。

3. 実験

本研究では、Tweet 日本語感情コーパスへの BERT の効果の検証のために以下の 2 つの実験で感情分類を行い、結果の比較と分析を行った。

実験1. バッチサイズ, ドロップアウト, 学習率の 3 つのパラメータについての分析を行い、各パラメータの関係を調査する。

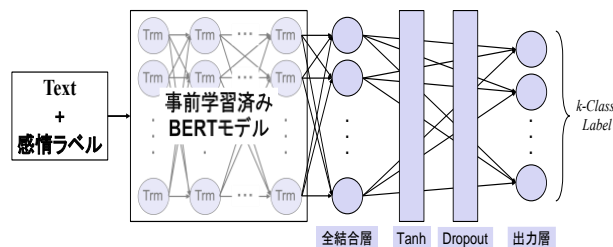


図 1 モデルの構造

表 1 データセット内の各感情の Tweet 数

	不安	怒り	悲しみ	嫌悪	喜び	愛	期待	驚き	無感情
Train	218	219	216	206	223	215	224	219	225
Test	100	100	100	100	100	100	100	100	100

実験2. BERT の微調整を行うことが、Tweet 日本語感情コーパスに有効かどうかの分析を行う。分析のために以下の 3 つの条件での F 値を比較する。

- 条件1. BERT の全ての層を微調整したモデル
- 条件2. BERT の最終層のみを微調整したモデル
- 条件3. BERT の微調整しないモデル

また、ベースラインとしてロジスティック回帰の F 値とも比較し、BERT の有効性を示す。

3.1. データセット

本研究では、訓練データを 1,600 文、テストデータを 600 文使用した。これらのデータセットは全て Tweet と感情ラベルで構成される。各感情の Tweet 数を表 1に示す。ここで、訓練データはあらかじめ収集しておいた Tweet を利用して能動学習により作成した。

本研究では、初期訓練データを 600 文とし、エントロピー計算, ユークリッド距離計算, カルバックライブラー距離計算により、テキストに表れる感情が曖昧で、文法が代表的であり、文法及び感情が多様なテキストの選択をする能動学習手法を用いる。また、テキストは常に感情ラベルの均衡性を保つように選択される。

Verification of the effect of BERT on the Tweet Japanese emotional corpus

[†]Tatsuya Ikegami, Graduate School of Science and Technology, Tokushima University

[‡]Fuji Ren, Shun Nishide, Xin Kang, Graduate School of Social, Industrial Science and Engineering, Tokushima University

表 2 最も F 値が高い組み合わせ

学習回数	ドロップアウト	バッチサイズ	学習率	Micro F1-score	Macro F1-score
5	0.1	16	4e-5	0.617	0.616

表 3 実験 2 の比較結果

	学習回数	Micro F1-score	Macro F1-score
ロジスティック回帰		0.528	0.531
条件 1.	5	0.617	0.616
条件 2.	10	0.597	0.598
条件 3.	62	0.546	0.547

3.2. 結果と考察

実験 1 より、バッチサイズが 8, 64 の場合に F 値が大きく下がる結果となった。しかし、ドロップアウトの値を上げることで改善された。これは、バッチサイズがデータセットに対して不適切であったため、過学習となったのではと考える。よって、バッチサイズとドロップアウトはデータの量に依存しているものといえる。また、学習率は他のパラメータに関係なく F 値が改善される値は全ての条件でほぼ同じであった。よって、学習率はデータの複雑さに関係しており、適切な値は分析するデータに対して一意に決まる可能性がある。F 値が最も優れていたパラメータの組み合わせを表 2 に示す。

実験 2 のそれぞれの条件下で F 値が収束した値と学習回数を表 3 に、各感情の F 値を表 4 に示す。本実験でのパラメータは、表 2 の通りである。結果より、微調整した BERT が最も F 値が高いことがわかる。一方、微調整なしの場合は、条件 1. と条件 2. に比べ F 値が少し低く、学習回数が増加している。よって、Tweet には、特有の特徴があり、それに対して BERT を微調整することが有益であることがわかる。また、表 3 より、条件 1. と条件 2. を比較した時、この 2 つは、ほぼ同等の結果を示していることがわかる。これは、

BERT の最終層が、微調整のための重要な層であるために得られた結果ではないかと考える。さらに、条件 2. は、条件 1. の約 80% の時間で F 値が収束する。この結果は、今後の BERT の学習時間削減の可能性を示唆する。

4. まとめ

本研究では、感情分類のために Tweet 日本語感情コーパスに対して BERT を利用した時の効果を検証した。結果より、微調整した BERT が最も有益であることが示された。また、BERT の学習時間の削減の可能性を示すことができた。一方で、若者言葉や Twitter の特有表現などの認識に関してはまだまだ改善する余地がある。

よって、今後の課題として、より特徴的な文脈を含むコーパスの作成ができる能動学習手法を検討し、感情分類精度の向上を目指す。

謝辞

本研究の一部は徳島大学研究クラスター事業(番号: 2003002)の支援を受けたものである。

参考文献

- [1]. Bian J, Zhao Y, Salloum RG, Guo Y, Wang M, Prospero M, Zhang H, Du X, RamirezDiaz LJ, He Z, Sun Y: "Using Social Media Data to Understand the Impact of Promotional Information on Laypeople's Discussions," A Case Study of Lynch Syndrome, J Med Internet Res 2017;19(12):e414, URL: <https://www.jmir.org/2017/12/e414>, DOI:10.2196/jmir.9266, PMID:29237586, PMCID:5745354
- [2]. J Devlin, M.-W. Chang, K. Lee, and K. Toutanova: "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," Proc. of NAACL-HLT 2019, pp. 4171-4186, 2019
- [3]. Kang Xin, Xuefeng Shi, Yunong Wu, and Fuji Ren: "Active learning with complementary sampling for instructing class-biased multi-label text emotion classification," IEEE Transactions on Affective Computing (2020)

表 4 各感情の F 値

	不安	怒り	悲しみ	嫌悪	喜び	愛	期待	驚き	無感情
ロジスティック回帰	0.493	0.554	0.490	0.523	0.694	0.468	0.586	0.464	0.505
条件 1.	0.609	0.635	0.595	0.585	0.707	0.548	0.643	0.589	0.639
条件 2.	0.560	0.626	0.539	0.616	0.720	0.538	0.652	0.555	0.581
条件 3.	0.498	0.583	0.551	0.571	0.663	0.446	0.587	0.471	0.555