

絵文字を用いて自動生成した感情コーパスの有効性について Effectiveness of Emotional Corpus Automatically Generated using Emoji

浅野 晴暉†
Haruki Asano

松原 雅文†
Masafumi Matsuhara

1. はじめに

近年、テキストから感情を機械的に識別する感情分析と呼ばれる技術の研究が盛んに行われている。一般に感情分析は、ネガティブ・ポジティブの二クラスを判別する手法が多く、喜・怒・哀・楽といった多クラスの感情を判別する手法は少ない。その理由としては利用可能な既存のデータセットが非常に少ないことがあげられる。

近年では、絵文字を用いることで上記の問題点を緩和し、多クラスの感情分析を実現している手法が存在する [1]。しかし、絵文字と感情の関係性の解釈が主観的であったり、あくまで疑似ラベルとして作用するためノイズが含まれてしまうといった問題点が存在する。

そこで、本稿では絵文字を用いて自動生成したコーパスの、感情分析における有効性を検証する。具体的には、絵文字予測による事前学習と Self-training を用いることにより、絵文字と感情の関係性を客観的に定義し、かつデータ内のノイズを緩和することで良質なコーパスを自動生成することを目指している。

2. 関連研究

多クラス感情分析では、感情語辞書を用いた手法や人為的にラベル付けした教師データで機械学習を行う手法が多数存在する。しかし、辞書ベースの手法では文法規則に対応できず、教師データの人為的なラベル付けは作成コストの点で限界があるなどの問題がある。そのため、近年では絵文字を活用した感情分析の研究が行われている [1]。

絵文字には感情を表す性質があるため、絵文字が表すと考えられる感情を疑似ラベルとすることで、絵文字を含むテキストに感情ラベルを自動付与することが可能となる。しかし、絵文字と感情の関係性の解釈が主観的であり、あくまで疑似ラベルとして作用するため、テキストと感情の関係が矛盾したノイズとなるデータが含まれてしまう場合がある。絵文字の感情を同定するのではなく、絵文字を予測する過程で得られた重みを感情分析モデルに転移することで分析精度を向上させた研究も存在する [2]。

本手法では、絵文字予測による事前学習と転移学習を行い、さらに Self-training を行う。これにより絵文字の感情値を定量的に把握し、かつノイズとなるデータを削減し、良質なコーパスを生成する。絵文字予測による事前学習と転移学習に関して、テキストのエンコード処理と転移学習による重みの調整方法が従来手法 [2] とは異なる。

3. 提案手法

3.1. 絵文字予測による事前学習

絵文字は多くの場合において感情を表す性質をもつため、あるテキストにどの絵文字が付与されているかを事前に予測することで、テキストと感情の対応を表した特徴量が算出されると考えられる [2]。処理としてはテキストをエンコードし得られた特徴量から絵文字のラベルを出力するように学習させる。

すべての種類の絵文字を予測対象とした場合、学習コーパスが増加し、かつ感情と関わりの浅い絵文字が含まれるといった問題点が存在する。そのため、感情を陽に表すと考えられる絵文字を抜粋しその絵文字のみを予測対象とする。具体的には Emojipedia¹ という絵文字リファレンス Web サイトを用い、予測する対象の感情をクエリとし検索した結果をもとに抜粋する。

3.2. 感情分析モデルの転移学習

事前に絵文字を予測したモデルを初期の重みとして活用し、感情分析モデルを学習する。教師データには WRIME [3] を用いる。このデータセットはテキストに対して Plutchik の 8 感情 (喜び, 悲しみ, 期待, 驚き, 怒り, 恐れ, 嫌悪, 信頼) を主観と客観両方の視点から 4 段階 (0:無・1:弱・2:中・3:強) でラベル付けしたものである。なお、本手法では Ekman の基本 6 感情 (喜び, 悲しみ, 驚き, 怒り, 恐れ, 嫌悪) を分析対象とするが、期待, 信頼は感情の識別が困難であると考えられるため除くものとする。

3.3. 絵文字の感情値算出

絵文字予測モデルと感情分析モデルより、総数 n のテキストに対し、 i 種類の絵文字の予測値を表す行列 E_{ni} と j 種類の感情の予測値を表す行列 S_{nj} が算出できる。これらの行列から絵文字と感情値の対応が分かり、絵文字の定量的な感情値が算出できると考えられる。そこで、絵文字の感情値を表す行列 V_{ij} を式 (1) により求める。

$$V = \frac{1}{n} E^T S \quad (1)$$

テキストから予測される絵文字の確率および感情値をかけ合わせ、 n 個の総和をとった値の平均値を絵文字の感情値として算出する。

3.4. Self-training による学習

Self-training は、ラベル付きデータ L で学習させた単一のモデルでラベルなしデータ U を予測し、確信度

†岩手県立大学ソフトウェア情報学研究科

¹<https://emojipedia.org/>

が高いデータ $x \in U$ に疑似ラベルを付与し L に加え、再学習を行う手法である。一般に識別モデルにおける確信度に用いるスコアは、入力 x に対する事後確率を用いることが多いため、性能によっては誤った疑似ラベルデータが生成され、不適切な学習が行われるリスクがある。

そこで、本手法では確信度に用いるスコアを式 (2) より求める。

$$\text{score}(x) = \lambda_1 p(C_i|x)V + \lambda_2 p(C_j|x) \quad (2)$$

$p(C_i|x)$ は入力 x に対する絵文字の予測値、 $p(C_j|x)$ は入力 x に対する感情の予測値を表す。 λ_1 、 λ_2 はパラメータであり、重みにより優先度を調整する。つまり、絵文字をもとに予測した感情値と感情分析モデルによる感情の予測値を両方用い、これらをパラメータで調整することで、ノイズとなるデータを削減し、良質なコーパスを生成する狙いがある。

4. 実験

4.1. 概要

絵文字予測による事前学習および感情分析モデルの転移学習に対する有効性を検証するため、テキストに対する感情分析精度を評価した。すなわち、提案手法において 3.2. 節までの処理に基づき学習を行い構築されたモデルの評価を行っており、絵文字に対する感情値の算出および Self-training による学習は行っていない。

事前学習の学習データには Twitter API を用いて収集した 10 万件の絵文字付きテキストを用いる。なお、対象とする絵文字を図 1 に示す。図 1 から確認できるように、今回は 32 種類の絵文字を対象とした。



図 1: 対象とした絵文字

テキストのエンコードには事前学習済みモデルの BERT²を用い、最適化アルゴリズムには AdaBound[4]を用いた。評価データには WRIME のテストデータ 2,000 件を用い、評価値には MAE を用い、モデルの出力した感情値とテストデータとの誤差を算出することで、テキストに対する感情分析の精度を評価する。

4.2. 実験結果

テキストに対する感情分析精度を表 1 に示す。表 1 より、MAE が 0.57 となっていることが確認できる。テストデータのラベル値の範囲が 0~3 であることを考慮すると、約 80% が正しく分析されているといえる。これは、絵文字の事前学習によりテキストと感情の潜在的な対応関係を学習し、高い精度で感情分析を行えた結果であると考えられる。しかし、今回は他の条件で

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

表 1: 感情分析精度

model	MAE
BERT+絵文字予測+転移学習	0.57

学習させたモデルとの比較を行っていない。そのため、今後は絵文字予測によって得られた重みの影響を強めた上での学習や、他の条件で学習させたモデルとの比較を兼ねて、絵文字予測による事前学習の有効性を再確認する必要がある。

また、今回は回帰での評価を行ったが、最も高い値のカテゴリは一致しているが、数値の誤差が大きくなっている予測値が多く見受けられた。そのため、今後は分類の評価も加えて考察を行う必要がある。

5. おわりに

本稿では、絵文字予測による事前学習と Self-training を用いた感情分析手法を提案し、絵文字を用いて自動生成した感情コーパスの有効性を検証した。実験結果から、絵文字予測による事前学習を行うことでテキストと感情の潜在的な対応関係を学習し、テキストに対する感情分析精度が向上することが確認された。

今後は、絵文字予測によって得られた重みの影響を強めた上での実験を行い、絵文字予測による事前学習の有効性を再確認する。また、絵文字の感情値算出および Self-training の実装を行い、提案手法により、最終的に良質な感情コーパスが生成できることを示す予定である。

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] 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.
- [2] 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. Association for Computational Linguistics.
- [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, Online, June 2021. Association for Computational Linguistics.
- [4] 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.