

# テキスト感情分析における絵文字疑似ラベル適用方法についての検討 Study on Methods of Applying Emoji Pseudo Label in Text Sentiment Analysis

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。感情分析は、ネガティブ・ポジティブの二クラスと、喜・怒・哀・楽といった多クラスの感情を分析する手法に大別されるが、特に後者の手法では高精度な分析が困難である。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。

そこで、絵文字を用いることで上記の問題を緩和し、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疑似ラベルを生成し、それを学習に適用する方法を検討する。特に、疑似ラベル内からノイズとなるデータを検知し、それを取り除く処理に焦点を当て、実験からその有効性を示す。

## 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、近年では絵文字から感情の疑似ラベルを生成し、それを感情分析に活用する研究が存在する [1]。絵文字は感情を表すことが多いため、絵文字と感情を同定し、その絵文字が含まれるテキストはその感情を示すと仮定し、ラベルを付与する方法である。しかし、上記の手法で生成された疑似ラベルデータにはノイズが含まれる。多量のデータを用意すれば、学習を行う中でノイズとなるデータの影響は少なくなるといった考えから、我々の研究では絵文字予測を補助タスクとしたマルチタスク学習を提案した [2]。しかし、伝達される知識の中からノイズとなる情報がどれだけ含まれているかを検証し、それをモデル内で除去することは容易ではない。そのため、疑似ラベルデータ内からノイズを直接取り除くことができる手法が必要であると考えられる。

そこで、本稿では絵文字をもとに生成した疑似ラベルデータ内からノイズとなるデータを自動検知し、それを取り除く手法を検討する。具体的には Self Attention

†岩手県立大学ソフトウェア情報学研究科

の値を用いてノイズを検知する。

## 3. 提案手法

### 3.1. モデルの設定

本手法では BERT [3] を用いる。テキストを入力とし、[CLS] に該当する特徴量を全結合しクラス分類を行うモデルを構築する。

### 3.2. 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ため、あるテキストにどの絵文字が付与されているかを予測することで、テキストと感情の適切なマッピングが得られると考えられる。予測値は式 (1)(2) により求められる。テキスト  $X = \{x_1, x_2, \dots, x_n\}$  を BERT に入力し、[CLS] に該当する特徴量  $C$  を得る。その後、 $C$  を重み  $W$  で全結合し softmax 関数にかけ、出力値を求める。

$$C = \text{BERT}(X) \quad (1)$$

$$y = \text{softmax}(CW + b) \quad (2)$$

### 3.3. Self Attention を用いた信頼度算出

式 (3) は、 $\sqrt{d_k}$  でスケールされた Dot-Product Attention である。

$$\text{Attention}(Q, K, V) = \text{softmax} \left( \frac{QK}{\sqrt{d_k}} \right) V \quad (3)$$

Self Attention は参照元の  $Q$  と参照先の  $K$  との内積を算出し、その値を重みとして特徴量を抜き出す演算である。そして、参照元と参照先が同一の入力となる。BERT の場合、 $C$  がテキスト特徴量として用いられるため、[CLS] を query として算出される各単語の Attention は、予測の寄与率のように解釈できる。一般的に、テキストの特徴を表す単語の Attention が高くなり、そういった単語が存在しない場合には [CLS] や [SEP] といった特殊トークンの Attention が高くなるといわれている [4]。そこで、特殊トークンの Attention が高いデータはノイズとなる割合が高いとみなし、Attention を信頼度のスコアに用いてデータのサンプリングを行う。

まず初めに絵文字予測モデルを学習し、学習したデータに対して予測を行う。そして、予測を行う際に [CLS] を query とした Attention を算出し、key が [CLS] と [SEP] 時の Attention を足し合わせたものをスコアとし算出する。そのスコアが一定の閾値よりも高い場合に、これをノイズとみなし取り除く。

## 4. 実験

### 4.1. 概要

Self Attention を用いたノイズ検知法の有効性を検証するために実験を行った。モデルには、日本語事前学習済みモデルの BERT<sup>1</sup> を用いる。本実験では以下のモデルを比較することで提案手法の有効性を示す。

- BERT<sub>BASE</sub> : すべてのデータを用いて学習
- BERT<sub>OUR</sub> : 提案手法に基づきノイズを除去したデータで学習

今回、閾値はすべてのデータのスコアに対する第三四分位数の値に設定する。

### 4.2. データセット

絵文字予測の学習データには Twitter API を用いて収集した、8 種類の絵文字が付与されたテキストを用いる。1 種類の絵文字に対して 2,000 件ずつ無作為にサンプリングした、総数 16,000 件のデータセットとなっている。対象とした絵文字を図 1 に示す。なお、絵文字と感情の同定は Emotag1200 を用いて行っている [5]。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 を用いる [6]。モデルの出力は絵文字であるが、同定した感情に変換することで評価を可能にしている。



図 1: 対象とした絵文字

### 4.3. 実験結果

各モデルの評価値を表 1 に記す。

表 1: 各モデルの感情カテゴリに対する MAE

モデル	悲しみ	驚き	怒り	平均
BERT <sub>BASE</sub>	0.33	0.29	0.15	0.29
BERT <sub>OUR</sub>	0.37	0.24	0.11	0.28

なお、分析は 8 種類の感情を対象にしているが、ここでは差が大きかった、悲しみ、驚き、怒りの結果を抜粋している。表 1 より、BERT<sub>BASE</sub> と BERT<sub>OUR</sub> 間で平均の MAE にほとんど差がなく、モデルの性能は同程度であることが確認できる。そのため、ノイズとして除去したデータは、もとより学習に与えている影響は少なかったと考えられる。しかし、提案手法を適用した際に、3,456 個のデータが除去されている。よって、少なくとも疑似ラベルデータ内から不必要なデータを

一定程度取り除くことができ、より少ないデータで同程度の性能を持つモデルを構築できたと考えられる。

また、感情のカテゴリによって、誤差の増減が異なるため、絵文字種類によってノイズの割合が異なっている可能性がある。よって、今後は定性分析を含めノイズの特徴などを整理する必要がある。

## 5. おわりに

本稿では絵文字をもとに生成した疑似ラベルデータ内からノイズとなるデータを自動検知し、それを取り除く手法を検討した。具体的には、Self Attention を用いてスコアリングしノイズを取り除く手法を提案し、実験により有効性を検証した。実験結果から、提案手法により、学習に不必要なデータを一定程度取り除くことができ、より少ないデータで同程度の性能を持つモデルを構築することができた。今後は、取り除かれたデータの定性分析や絵文字の割合を算出することで、ノイズと判断されたデータの特徴を分析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] 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] 浅野晴暉, 松原雅文. テキスト感情分析におけるマルチタスク学習の有効性. 情報処理学会第 84 回全国大会, pp. 1V–06, March 2022.
- [3] 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*.
- [4] Kevin Clark, Urvashi Khandelwal, Omer Levy, and Christopher D. Manning. What does BERT look at? an analysis of BERT's attention. In *Proceedings of the 2019 ACL Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*.
- [5] 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)*.
- [6] 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*.

<sup>1</sup><https://huggingface.co/cl-tohoku/bert-base-japanese-whole-word-masking>