

BERTによる商品レビューの観点付き感情極性分析

Aspect-based sentiment analysis in product reviews by BERT

板垣 紫音
Shion Itagaki

岡山大学 太田研究室
Ohta Laboratory, Okayama University

概要 本研究では、商品の評判を観点ごとに可視化するため、BERT(Bidirectional Encoder Representations from Transformers)による商品レビューの観点付き感情極性分析を行う。具体的にはまず Amazon の掃除機の商品レビューを用いたデータで事前学習済みの BERT をファインチューニングする。ファインチューニングに用いるデータは、観点とその観点に関する感情がポジティブ（正）なものかネガティブ（負）なものかを合わせたラベルが付いたレビューである。そして、その BERT でテストデータのレビューの観点とその感情極性を分類する。実験の結果、“吸引力：正”、“重量：正”、“吸引力：負”では再現率、適合率、F 値がいずれも 0.5 以上で、他のクラスより相対的に良かったが、デザインの観点に関しては全く分類できなかった。

1 はじめに

近年 EC サイトが普及し、インターネット上で商品を購入する機会が増えている。EC サイトのユーザは他のユーザの商品レビューを元に商品と比較し、購入することが多い。よって、多くのレビューが存在するため、そのすべての内容を把握し、商品の観点ごとの評判を理解することは困難である。しかし、レビューの観点ごとの評判を可視化することができれば、ユーザの商品の比較、購入は容易になる。観点ごとの評判を可視化するにはレビュー文の観点付き感情極性分析が有用である。

中山ら[1]は日本語を対象とした観点付き感情極性分析のためのコーパスを構築した。楽天トラベルのユーザレビューに対し、“サービス”、“風呂”、“立地”、“朝食”、“設備・アメニティ”、“部屋”、“夕食”の 7 種類の観点についてポジティブかネガティブかの極性のタグを付与し、BERT によってマルチラベル分類を行った。

藤本ら[2]は評判情報を用いて代替品を推薦する手法を提案した。提案手法では、まず Google Natural Language API によって商品レビューのテキスト内の単語の感情極性を表す感情スコアを求めた。そして、商品の特徴を表す単語をまとめた単語クラスタを、word2vec によって得た単語の分散表現と k-means 法によるクラスタリングによって求めた。また商品レビューを可視化させるために、その単語クラスタをレーダーチャートの特徴軸とした。特徴軸の単語クラスタに含まれる単語の感情スコアの平均を特徴軸の感情スコアとし、その感情スコアの類似度に基づいて、代替品を推薦した。

本稿では、商品の観点ごとの評判を得るために、掃除機の商品レビューを用いたデータで BERT をファインチューニングする。なお、ファインチューニングには中山らの方

表 1：タグとタグが付与された文の数

タグ	タグが付与された文の数
吸引力：正	261
重量：正	122
音：正	27
デザイン：正	17
吸引力：負	78
重量：負	40
音：負	25
デザイン：負	11

法[1]を参考に、レビューに観点とその極性を表すタグを付けたデータを用いる。そして、その BERT でレビューを観点とその極性を表すクラスに分類し、その分類精度を検証した。

本稿では、まず 2 節で商品レビューの観点と極性について説明する。3 節で商品レビューの BERT による分類実験について説明する。最後に 4 節でまとめる。

2 商品レビューの観点と極性

商品レビューの観点付き感情極性分析をするため、Amazon で販売されているアイリスオーヤマ製の掃除機(型番 SCA-110-W)の 422 レビューを人手で取得する。取得したレビューを句点や感嘆符等、文末として判断できるものを手掛かりに 1 文ずつに分割する。レビューをすべて分割すると、合計 1660 文となった。その 1660 文に観点とその極性を表すタグを付与する。

本稿では掃除機のレビューを対象とするため“吸引力”、“重量”、“音”、“デザイン”の 4 つを観点とした。

この 4 種類の観点が含まれるレビューについて、その観点的情報がポジティブ（正）かネガティブ（負）かを著者が判定し、“吸引力：正”や“吸引力：負”のような合計 8 種類のタグを 1 つ以上付与する。その結果 1660 文の内 503 文にタグが付与できた。表 1 にタグが付与された文の数をタグごとに示す。なお、この文の数には重複が含まれる。

以下に Amazon のアイリスオーヤマの掃除機のレビューの分割された文と付与したタグの例を示す。

- まず皆さんが知りたいであろう「吸引力」これは完全にへたな高価掃除機よりあります。(吸引力：正)
- 軽いので使い勝手は良い。(重量：正)
- 音も比較的静かで良いと思います。(音：正)
- 見た目に関しても安っぽさはあまりなく、値段の割には良いほうだと思います。(デザイン：正)

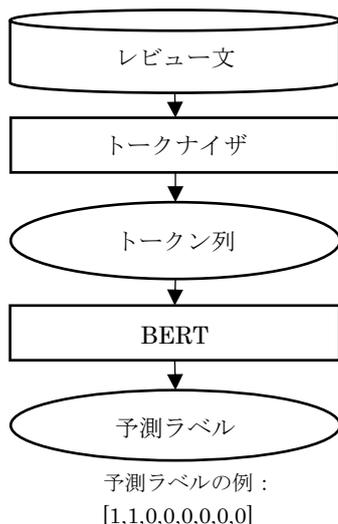


図 1: 商品レビューの分類の流れ

3 商品レビューの BERT による分類実験

3.1 BERT による商品レビューの分類

観点とその極性のタグが付けられた掃除機の商品レビュー一文で BERT をファインチューニングする。その BERT で掃除機についての商品レビューを観点と極性で表わされるクラスに分類する。

BERT は 2018 年 10 月に Google の Jacob Devlin らが発表した自然言語処理モデルである[3]。このモデルは様々な自然言語処理タスクにおいて高い精度を達成した。また、BERT はファインチューニングを行うことで様々なタスクに応用でき、高い汎用性を持つ。

掃除機の商品レビューの BERT による分類処理の流れを図 1 に示す。まずレビューをトークナイザによってトークンに分割する。つづいて分割されたトークンを BERT に入力する。そして、最後に出力として 8 要素を持つリストが出力される。このリストの要素の値が 1 となるとそのクラスに分類されたことになる。分類結果は 8 種類のクラスのうち 1 つ以上のクラスへの分類か、もしくはどのクラスにも分類されないかである。

BERT の日本語事前学習モデルには、Huggingface の `trancformers` で公開されている東北大学が学習したモデル¹を使用する。トークナイザもこの事前学習モデルで使われている `BertJapaneseTokenizer` を使用する。出力層の活性化関数としてシグモイド関数、損失関数として Binary Cross Entropy、最適化関数として Adam を用いている。

3.2 実験環境

実験に用いるデータは 2 節で説明した観点とその感情極性が付与された 503 文とする。503 文の内の 307 文を訓練用データ、98 文を検証用データとして BERT をファインチューニングする。バッチサイズは 32、エポック数は 16 とし

表 2: 掃除機のレビューの分類結果

	TP	FN	FP	TN	再現率	適合率	F 値
吸引力: 正	36	11	10	41	0.77	0.78	0.77
重量: 正	22	6	4	66	0.79	0.85	0.77
音: 正	1	3	2	92	0.25	0.33	0.28
デザイン: 正	0	6	0	92	0.00	0.00	0.00
吸引力: 負	9	7	8	74	0.56	0.53	0.54
重量: 負	2	5	4	87	0.40	0.50	0.44
音: 負	3	4	0	91	0.43	1.00	0.60
デザイン: 負	0	2	0	96	0.00	0.00	0.00

た。分類精度は残りの 98 文のテストデータを使用して算出する。

3.3 実験結果

掃除機のレビューの観点付き感情極性分類の結果は表 2 のようになった。表 2 では観点と極性のクラスを正しく予測したものが True Positive (TP)、正しくないのにそのクラスと予測したものが False Positive (FP)、誤ってそのクラスと予測できなかったものが False Negative (FN)、正しくそのクラスでないで予測したものが True Negative (TN) である。また、再現率、適合率、F 値は次式で計算される。

$$\text{再現率} = \frac{TP}{TP + FN} \quad (1)$$

$$\text{適合率} = \frac{TP}{TP + FP} \quad (2)$$

$$\text{F 値} = \frac{2 * \text{再現率} * \text{適合率}}{\text{再現率} + \text{適合率}} \quad (3)$$

3.3 考察

表 2 から“吸引力: 正”と“重量: 正”と“吸引力: 負”のクラスについては再現率、適合率、F 値がいずれも 0.5 以上あり、他のクラスよりも相対的に分類精度が高かった。

表 2 から全体的に再現率よりも適合率の方が高く、取りこぼしが多いが、レビューに間違ったラベルを付けにくくなっている。

4 まとめ

本稿では観点付き感情極性のタグ付けをした掃除機に関するレビューでファインチューニングした BERT を用いて、レビュー分類を行った。レビューには“吸引力”、“重量”、“音”、“デザイン”の 4 種類の観点と正か負の極性を表す 8 種類のタグが 1 つ以上付与されている。

“吸引力: 正”、“重量: 正”、“吸引力: 負”のクラスについては比較的分類ができたが、レビュー数の少なかったデザインに関しては全く分類ができなかった。また、

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

取りこぼしは多いが、間違ったクラスには分類されにくかった。

今後の課題として、レビューが少なく分類精度が低かったクラスの分類精度向上のために、そのクラスのレビューデータをさらに取得することや、商品レビューをまず観点に分類し、その後感情極性を分析する方法などを検討する予定である。

参考文献

- [1] 中山祐輝, 村上浩司, Ikuko Hardaway, “アスペクトベース意見分析における日本語評価コーパス”, 言語処理学会第 27 回年次大会発表論文集, pp. 375–380, 2021.
- [2] 藤本大吾, 上野史, 太田学, “評判情報を用いた代替品推薦の一手法”, DEIM2022, C34-2, 2022.
- [3] Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”, arXiv:1810.04805, 2018.