

日本語 BERT モデルを用いたレビュアーの心情抽出

中田聖之[†]
愛知県立大学[†]山村毅[‡]
愛知県立大学[‡]

1 はじめに

近年、商品に対するレビューを投稿することができるネット通販サイトが多数存在する。商品レビューには購入者の感想や意見が込められているので、ユーザや企業にとって有益な情報である。しかし、膨大な商品レビュー群から自分の知りたい情報を発見するのに時間を要することもある。よって、利用者が欲する商品レビューを効率的に提供する技術が、より快適なネットショッピングを実現する上で重要であるだろう。そこで、商品レビュー文に込められている投稿者の意見のタイプ（以下、心情）を特定することで、肯定的・否定的な商品レビュー、製作者への提言、購入検討者へのアドバイスなどの絞り込みに貢献し、キーワードや星の数による従来の絞り込みよりも高度な商品レビューの抽出が実現できると考えた。本研究では、東北大学 乾研究室の事前学習済み日本語 BERT モデル（以下、BERT）[1] をはじめとする様々な条件の機械学習モデルで心情抽出を行う実験を実施する。そして、実験結果を比較することで、BERT の各機構が心情抽出の精度向上のためにどのような役割を果たしているのかを明らかにする。

2 関連研究

藤巻ら [2] は、毎日新聞の社説に対して Bag of Words による特徴表現を施し、ナイーブベイズ分類器と最大エントロピー法で心情抽出を行った。そして、Bag of Words を用いた場合の分類器の違いが心情の抽出に与える影響について調べた。宮川ら [3] は、毎日新聞の社説に対して表記ゆれを考慮する場合としない場合に対し、word2vec で心情抽出を行いその精度を比較することで、表記ゆれの有無が心情の抽出に与える影響について調べた。

3 使用するカテゴリとデータ

本研究では、レビュアーの心情が表 1 のように分類できると仮定する。本研究で使用する商品レビュー文は、「Amazon.com」において収集を行った、5000 件のレビューである。この商品レビュー文は、特定のジャンルに偏らないように収集した。商品レビュー文の学習・検証と、テストに使用するデータの件数および心情のカテゴリの内訳は、表 2 の通りである。

4 機械学習モデル

本研究で扱う機械学習モデルの名称と説明を以下に示す。

表 1 カテゴリの一覧

対象	感情	カテゴリ	説明
商品	肯定的	肯定	商品に関する肯定的な意見
	否定的	否定	商品に関する否定的な意見
製作者・出品者		期待	製作者に向けた要求・提言
購入検討者		助言	購入者に向けた助言
その他		その他	上記以外のもの

表 2 商品レビュー文の内訳

役割	カテゴリ					合計
	肯定	否定	期待	助言	その他	
学習・検証	750	750	750	750	750	3750
テスト	250	250	250	250	250	1250
合計	1000	1000	1000	1000	1000	5000

BERT FC BERT のプールされた出力に線形層を備えたニューラルネットワーク。

BERT MFC BERT のプールされた出力に 2 層の線形層を備えたニューラルネットワーク。これと BERT FC の性能を比較することで、分類器の違いによる精度の違いを調べる。

BERT SVM BERT で特徴表現したベクトルから SVM で多クラス分類する機械学習モデル。BERT に入力を与えるのは特徴表現のときのみであり、BERT のファインチューニングは行わない。これと後述する BoW SVM の性能を比較することで、BERT による特徴表現と Bag of Words による特徴表現での精度の違いを調べる。

EMB FC Embedding Bag と、線形層によるニューラルネットワーク。これと BERT FC の性能を比較することで、BERT による特徴表現と Embedding Bag による特徴表現での精度の違いを調べる。

BoW SVM Bag of Words で特徴表現し、SVM で多クラス分類をする機械学習モデル。

BoW FC Bag of Words で特徴表現し、線形層で多クラス分類をするニューラルネットワーク。これと BoW SVM の性能を比較することで、線形層と SVM の精度の違いを調べる。

5 評価方法

二つのシステム A,B に共通の N 個のデータを与えた結果、これらのシステムの出力の違いが有意なものかを検定するためにマクネマー検定を用いる。表 3 が与えられたとき、対応するラベルを持つセルでの発生する確率を p_i ($i = a, b, c, d$) とすると、帰無仮説 H_0 と対立仮説 H_1 は、 $H_0 : p_b = p_c$ vs. $H_1 : p_b \neq p_c$ となる。このとき、マクネマー検定では、統計量 $(b-c)^2/(b+c)$ が H_0 の下で χ^2 分布に従うことを利用して有意差検定を行う。

Feeling extraction of reviewers using the Japanese BERT model

[†] NAKATA Masayuki, Aichi Prefectural University

[‡] YAMAMURA Tsuyoshi, Aichi Prefectural University

表3 システム A × システム B のクロス集計表

		システム B		
		正解	不正解	計
システム A	正解	<i>a</i>	<i>b</i>	<i>a + b</i>
	不正解	<i>c</i>	<i>d</i>	<i>c + d</i>
	計	<i>a + c</i>	<i>b + d</i>	<i>N</i>

6 実験結果

各機械学習モデルの正解率および有意水準 5% のマクネマー検定による有意差判定の結果を表 4 に示す。

表4 各モデルの正解率と有意差検定

モデル名	正解率 [%]	有意差判定		
		比較対象	P 値 [%]	結果
BERT FC	90.88	—	—	—
BERT MFC	91.68	BERT FC	17.36	× ¹
BERT SVM	80.24	BERT FC	1.7e-17	○ ²
		BoW SVM	16.03	×
EMB FC	79.44	BERT FC	3e-18	○
BoW SVM	78.24	BERT FC	5e-20	○
BoW FC	77.84	BERT FC	2e-22	○
		BoW SVM	8.84	×

¹ ×: 有意差なし (有意水準 5% で帰無仮説が受容)

² ○: 有意差あり (有意水準 5% で帰無仮説が棄却)

7 考察

7.1 機械学習モデルに関する考察

表 4 の実験結果から、機械学習モデルに関する考察を行う。BERT FC と BERT MFC の精度に有意差は無い。つまり、BERT の分類器において、線形層が 2 層になることの有意差は無い。Bag of Words と BERT の特徴表現の間には有意差は無い。有意水準 5% で EMB FC より BERT FC の方が高精度であるので、心情抽出において、分散表現 (Embedding Bag) より BERT の特徴表現の方が優れている。線形層と SVM の間には有意差が無いので、心情抽出において、SVM と線形層という 2 つの分類器の性能の違いには有意差がない。つまり、BERT SVM の精度が BERT より劣る原因は Transformer Encoder のファインチューニングを行わないことのみに起因する。以上より、BERT を使わないモデルの正解率は高々 80% である。しかし、BERT にファインチューニングを施してモデルを作成する BERT FC と BERT MFC の正解率は 90% 以上である。このことから、心情抽出において BERT は他のモデルよりも優れていることがわかった。

7.2 レビュー文に関する考察

テストデータに用いた商品レビュー文と実験結果から、BERT による心情抽出の特徴について考察する。BERT を心情抽出に用いることの利点は二つある。

一つ目は、「文脈理解」の精度が向上するということである。心情の決定には、「心情を示す単語」(例えば、『肯定』なら「最高」、『否定』なら「残念」、『期待』なら「欲しい」、『助言』なら「オススメ」) が含まれているかどうかが大きく影響することがある。全てのモデルで正解だったレビュー文を調べてみる

と、「心情を示す単語」を含んだ文の割合は 12.2% であった。一方、BERT FC と BERT MFC の場合は、その割合は 3.6% にまで低下した。すなわち、BERT FC や BERT MFC は、「心情を示す単語」以外を含んだより多くのレビュー文に対して、正しく心情推定できたことになる。このことは、Transformer Encoder のファインチューニングをすることで、「心情を示す単語」によらない、文脈を考慮して心情を読み取る「文脈理解」の精度が向上することを示唆している。

二つ目は、「語順」を考慮することができるということである。通常、逆接 (が・のに・しかし) を含んだ文では、主張点は逆接の後ろにあることが多い。逆接を含む文が占める割合は、テストデータ全体では 19.2% であった。一方、BERT FC と BERT MFC と BERT SVM だけが正解だったレビュー文について調べてみると、逆接を含む文の割合は 35.4% に増えていた。すなわち、BERT を用いたモデルでは、逆接を含んだ文のより多くを正しく心情推定していた。このことは、「語順」を考慮した心情抽出をしていることを示唆している。これは、BERT が Positional Embedding を用いて単語の位置情報を特徴量ベクトルに与えていることが影響していると考えられる。

文脈理解に強い BERT をもってしても、特定の心情を表す単語に影響を受けずに言葉の裏を読んで心情を抽出することは難しい。具体的には、「ファンの方も、そうでない方も大いに満足できる最高のベストアルバムだと思います。」というレビュー文がある。この文は、「最高」という『肯定』を連想させる単語が含まれるが、購入検討者に向けてファンでもファン以外でも満足できる商品であることをアドバイスする『助言』である。このように、特定の心情を示す単語を含むが、実際のカテゴリはその単語が示すカテゴリとは異なるようなレビュー文は、本研究で扱ったすべての機械学習モデルで不正解だった。

8 まとめ

本研究では、東北大学乾研究室の事前学習済み日本語 BERT モデルをはじめとする様々な条件の機械学習モデルに対して、Amazon のレビュー文からレビュー者の心情抽出を行う実験を実施した。実験の結果、BERT を心情抽出に用いることの利点は二つあることがわかった。文脈から心情を読み取れるようになることと、語順を考慮できるようになることである。一方、BERT でも言葉の裏を読んで心情を抽出することはできない。今後は、言葉の裏を読んで心情抽出をすることができるような手法を考えていきたい。

参考文献

- [1] 乾健太郎:『日本語 BERT モデル公開』, 東北大学 情報科学研究科 自然言語処理学, 2019.
<https://www.nlp.ecei.tohoku.ac.jp/news-release/3284/> (最終閲覧日: 2022/12/09)
- [2] 藤巻直也, 三浦弦太, 山村毅:『意見情報の推定における Bag-of-Words の有効性について』, 電気・電子・情報関係学会東海支部連合大会講演論文集, D4-3, 2015.
- [3] 宮川知也, 山村毅:『感性情報分類における見出し語の違いについて』, 自然言語処理学会第 23 回年次大会発表論文集, 2017.