

衣服を対象とした商品レビューからの長所・短所の抽出

井上京¹ 山田剛一² 増田英孝²

東京電機大学大学院未来科学研究科¹ 東京電機大学未来科学部²

22fmi06@ms.dendai.ac.jp {yamada,masuda}@mail.dendai.ac.jp

概要

近年、アパレル通販市場が拡大しており、商品レビューはユーザが商品を購入するかどうか決める際の大きな手掛かりとなっている。しかし、商品レビュー数が多い場合、それらをすべて読むことは難しい。そこで、本研究では商品の長所・短所を提示することを目的とし、商品レビューから商品の長所・短所を抽出する手法を提案する。商品レビューから属性と評価値を抽出し、それらに対する感情の3つをトリプレットとして予測する。また、商品レビューに属性が明記されていない場合には、その属性の推定も行う。評価実験において、トリプレット抽出のF1スコアが「Tシャツ・カットソー」カテゴリで0.54、「セーター」カテゴリで0.53であった。

1 はじめに

近年、コロナ禍の影響でオンラインの需要が伸びたことによりアパレル通販市場が拡大しつつある[1]。また、多くの通販では商品レビューを投稿する機能があり、ユーザは他のユーザが投稿した商品レビューを見ることができる。商品レビューは実際に商品を購入した人の意見や感想が記載されており、これらはユーザが商品を購入するかどうかを決める際の大きな手掛かりとなっている。

しかし、商品レビュー数が多い場合、それらをすべて読むのに膨大な時間がかかってしまう。また、一部を読んだだけでは商品の長所や短所をすべて把握することは難しい。

そこで、本研究ではユーザが商品レビューをすべて読まなくても商品の長所や短所を把握できるシステムを作ることを目的とし、商品レビューから商品の長所・短所を抽出する手法を提案する。商品レビューから属性と評価値を抽出し、それらに対する感情を予測する。最終的には、これら3つをまとめてトリプレットとして抽出する。また、商品レビューに明記されていない場合には、その属性の推定も行う、より網羅的な抽出を図る。

2 関連研究

Wu ら[2]はレビューからトリプレットを抽出することを目的とした、グリッドタギング方式(GTS)を提案している。GTSは6種のタグ{A, O, Pos, Neu, Neg, N}を用いて文中の任意の2つの単語関係を表すものである。CNN, BiLSTM, BERT をエンコーダとしてそれぞれ学習モデルを構築し、検証した結果、BERT が最も良い精度を得た。

Mukherjee ら[3]はタグ付けを用いない手法として、ポインタネットワーク(PASTE)を提案している。PASTE では属性の開始位置と終了位置、評価値の開始位置と終了位置、感情を予測することにし、位置に基づいた予測を行うことにより、属性と評価値の単語の長さが異なってもトリプレットを一様に表現することができる。

上記を含む多くの先行研究は属性が文中に明記されていることを前提とした手法を提案しており、これらが文中に明記されていない場合の考慮はされていない。しかし、これらが文中に明記されていない場合の抽出手法を提案している先行研究も存在する。Cai ら[4]はトリプレットにカテゴリを加えたクアドラブルの抽出を提案している。属性が文中に明記されていない場合は属性を null で抽出し、その分をカテゴリで補う。しかし、最も良い結果で F1 スコア 0.446 であり、改善の余地がある。柏木ら[5]は省略された属性の推定も含めた属性-評価値ペア抽出を提案している。事前に属性の候補リストを作成し、この中から省略された属性を推定する。また、レビュー文中の属性・評価値を抽出するための属性辞書・評価値辞書を作成している。文書中に明記されていない属性の抽出は F 値 0.55、文書中に明記されている属性の抽出は F 値 0.24 だった。

日本語の商品レビューは、基本的に評価値は文中に明記されているが、属性は明記されていない場合が多くある。そのため、本研究では、属性が文中に明記されているトリプレット抽出と属性が文中に明記されていないトリプレット抽出を同時に行う。



図 1 アノテーションの例(属性有り)

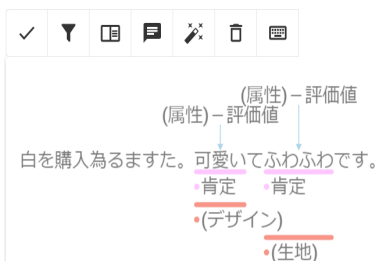


図 2 アノテーションの例(属性無し)

3 提案手法

本章では、本研究で使用するデータセットや前処理、トリプレット抽出の手法について述べる。

3.1 用語の定義

本研究で扱う用語を以下のように定義する。

- 属性：評価対象となる商品の評価視点
- 評価値：属性に対する評価
- トリプレット：属性，評価値，感情をまとめて抽出したもの

3.2 データセット

本研究では、楽天グループ株式会社が提供している「楽天データセット」[6]の商品レビューデータを使用し、その中から「T シャツ・カットソー」カテゴリと「セーター」カテゴリのデータを使用する。上記のカテゴリの特徴とこれらを使用する理由は以下の通りである。

- T シャツ・カットソー：基本的に1年中需要がある
- セーター：基本的に秋冬に需要がある
- 本研究で使用する理由：どの季節にも需要があるカテゴリと偏った季節に需要があるカテゴリの比較を行うため

なお、このデータセットには属性，評価値，感情のアノテーションが付与されていないため、本研究の手法に対応するためにアノテーションの付与を行

	The	hot	dogs	are	top	notch	but	average	coffee	
N	N	N	N	N	N	N	N	N	N	The
	A	A	N	Pos	Pos	N	N	N	N	hot
		A	N	Pos	Pos	N	N	N	N	dogs
			N	N	N	N	N	N	N	are
				O	O	N	N	N	N	top
					O	N	N	N	N	notch
						N	N	N	N	but
							O	Neu	N	average
								A	N	coffee

図 3 GTS のタグ付け結果例([2]から引用)

表 1 各タグの意味

タグ	意味
A	単語ペア(wi, wj)の2語が同じ属性に属している。
O	単語ペア(wi, wj)の2語が同じ評価値に属している。
Pos	単語ペア(wi, wj)の2語がそれぞれ属性と評価値に属し、単語ペアの感情が肯定である。
Neu	単語ペア(wi, wj)の2語がそれぞれ属性と評価値に属し、単語ペアの感情が中立である。
Neg	単語ペア(wi, wj)の2語がそれぞれ属性と評価値に属し、単語ペアの感情が否定である。
N	上記のいずれにも該当しない。

う。

3.2.1 前処理

アノテーションの付与を行う前にデータの前処理を行った。これを行う理由として、表記ゆれによる機械学習精度の低下が挙げられる。例えば、「あたたかい」と「暖かい」は本質的には同じ単語であるが、機械学習モデルはこれらを同一であると判断することができない。そのため、事前に表記ゆれの統一を行う必要がある。具体的には SudachiPy と Unicode 正規化を用いてデータに収録されている商品レビュー文の正規化を行い、表記ゆれの統一を行う。

3.2.2 アノテーション

前処理を行った商品レビュー文に対し、アノテ

ションの付与を行う。アノテーションの例を図 1、図 2 に示す（アノテーション対象のテキストは活用する語の語形を基本形に変換している）。今回は各カテゴリ約 1,500 件のデータに対し、アノテーションを行い、学習データ、検証データ、テストデータをおよそ 7:2:1 の割合で分割する。

3.3 トリプレットの抽出

本研究では、2 節で紹介した GTS[2]の手法を日本語のデータでも適応できるようにする。また、属性が文中に明記されていないトリプレットの抽出にも適応できるようにする。

3.3.1 GTS

GTS は 6 つのタグ {A, O, Pos, Neu, Neg, N} を用いて文中の任意の 2 つの単語関係を表すものである。タグ付け結果の一例を図 3 に示す。また、6 つのタグの意味を表 1 に示す。なお、Pos, Neu, Neg がタグ付けされた場合、その単語ペア (wi, wj) は属性-評価値ペアであることを表している。また、本研究では Pos を肯定、Neu を中立、Neg を否定と言い換えて感情を予測する。

3.3.2 日本語データへの適応

まず、McCab[7]を用いてレビュー文の分かち書きを行う。英語の文は単語ごとに空白があるため、文を容易に分割することができるが、日本語の文は各単語が連結されているため、事前に分かち書きをする必要がある。また、トークナイザを BertTokenizer から BertJapaneseTokenizer に変更する。これにより、文を BERT モデルに入力できる形にする際、より日本語に適した形に変換することができる。

3.3.3 属性が文中に明記されていないトリプレットの抽出

文中に明記されていない属性を推定するための属性リストを作成する。属性リストに載っている単語を表 2 に示す。

表 2 属性リストの単語一覧

デザイン	生地
着心地	価格
サイズ	商品
長さ	重さ

属性が文中に明記されている場合は属性と評価値を同時に抽出していたが、属性が文中に明記されて

いない場合は先に評価値のみ抽出する。次に、評価値に対応する属性を属性リストの中から選出し、これらを仮の属性-評価値ペアとする。その後、仮の属性-評価値ペアの感情を予測することで、属性が文中に明記されていない場合でもトリプレットを抽出できる。

4 実験

本章では、トリプレットの抽出精度を確認するため、3.2 節で作成したデータセットを用いてトリプレットの抽出実験を行う。今回は以下の 3 パターンで実験を行う。

- 属性が文中に明記されているトリプレットのみ抽出
- 属性が文中に明記されていないトリプレットのみ抽出
- 上記 2 種類のトリプレットを同時に抽出

なお、学習データ、検証データ、テストデータは各パターンに該当するトリプレットのみ使用する。各パターンのデータの詳細を表 3, 表 4, 表 5 に示す。

表 3 データの詳細(属性有り)

カテゴリ	レビュー文数	トリプレット数
T シャツ・学習	938	2,164
カットソー 検証	140	304
テスト	271	657
セーター 学習	1,017	2,531
検証	144	334
テスト	316	826

表 4 データの詳細(属性無し)

カテゴリ	レビュー文数	トリプレット数
T シャツ・学習	475	628
カットソー 検証	65	91
テスト	132	178
セーター 学習	506	706
検証	75	112
テスト	145	198

表 5 データの詳細(両方)

カテゴリ	レビュー文数	トリプレット数
T シャツ・カットソー	学習	2,792
	検証	395
	テスト	835
セーター	学習	3,237
	検証	446
	テスト	1,024

4.1 実験設定

事前学習済み BERT は東北大学が提供している bert-base-japanese-v3¹を使用する。また、学習率を $5e-5$ 、バッチサイズを 4、ドロップアウト確率を 0.1 に設定した。モデルパラメータの最適化には AdamW を用いた。実験では、各データ 30 エポックの学習を行う。

4.2 実験結果

各データの実験結果をそれぞれ表 6, 表 7 に示す。なお、トリプレットは属性、評価値、感情すべてが正しく予測された場合のみ正解とする。

表 6 実験結果 (T シャツ・カットソー)

パターン	適合率	再現率	F1 スコア
属性有り	0.641	0.531	0.581
属性無し	0.647	0.563	0.602
両方	0.585	0.501	0.540

表 7 実験結果 (セーター)

パターン	適合率	再現率	F1 スコア
属性有り	0.599	0.447	0.512
属性無し	0.573	0.638	0.604
両方	0.540	0.519	0.529

5 考察

両データ共に F1 スコアで 0.53~0.54 の値となった。また、今回の実験結果は Wu ら[2]が行った英語データでの評価実験の結果との間に大きな差がなかった。このことから、GTS[2]は日本語のデータへの適応も可能であることがわかった。

パターンごとの結果を見ると、属性無しトリプレ

ットのみの抽出が最も良い結果となった。これは属性の選択肢が 8 個のみだったため、属性の誤推定が少なかったことが要因だと考えられる。また、属性有りトリプレット、属性無しトリプレットを同時に抽出すると別々に抽出するよりも精度が低い結果となった。同時に抽出する時に使用したデータは、属性有りデータと属性無しデータを混ぜて作成しているため、別々に抽出する時よりも学習の難易度が上がっていることが原因の 1 つだと考えられる。

また、セーターにおける属性有りトリプレット抽出の再現率が他の項目と比べて著しく低い結果となった。これは 1 回しかデータセットに出現していない属性・評価値の数が多かったことが原因の 1 つだと考えられる。この課題を解決するには、データ数を増やすことが必須である。なお、属性・評価値の出現回数の詳細は付録 A に記載する。

全体的な結果としては、セーターの方が T シャツ・カットソーより精度が低い結果となった。これはセーターの方が T シャツ・カットソーより属性、評価値の種類が共に多く、その分属性-評価値ペアの組み合わせ数も多くなることから、誤った組み合わせで抽出してしまった可能性がある。

6 おわりに

本研究では、ユーザが商品レビューをすべて読まなくても衣服の長所や短所を把握できるシステムを実装することを目的とした、商品レビューからの衣服の長所・短所抽出手法を提案した。既存の手法を日本語データへ適応できるようにし、属性が文中に明記されていないトリプレットの抽出を可能とした。評価実験を行った結果、F1 スコアで 0.53~0.54 の値となった。今後の課題として、データ数を増加することや他の手法の検討が挙げられる。また、トップだけでなく、ボトムスなどの他カテゴリのデータを用いた評価実験も行いたい。

¹ <https://huggingface.co/cl-tohoku/bert-base-japanese-v3>

謝辞

本研究では、国立情報学研究所の IDR データセット提供サービスにより、楽天グループ株式会社から提供された「楽天データセット」を利用した。ここに記して感謝する。

参考文献

- [1] Commerce Marketing Blog : アパレル業界を目指す方が読むべき市場規模や平均年収まとめ！, <https://www.interfactory.co.jp/blog/apparel/>, 12月28日参照
- [2] Z.Wu, C.Ying, Z.Fan, et al: Grid Tagging Scheme for Aspect-oriented Fine-grained Opinion Extraction, In *Findings of the Association for Computational Linguistics EMNLP 2020*, pp.2576-2585(2020).
- [3] R.Mukherjee, T.Nayak, Y.Butala, et al: PASTE A Tagging-Free Decoding Framework Using Pointer Networks for Aspect Sentiment Triplet Extraction, In *Proceeding of the 2021 Conference on Empirical Methods in Natural Language Processing*, pp.9279-9291(2021).
- [4] H.Cai, R.Xia, J.Yu: Aspect-Category-Opinion-Sentiment Quadruple Extraction with Implicit Aspects and Opinions, In *Proceeding of the 59th Annual Meeting of the Association for Computational Linguistics and 11th International Joint Conference on Natural Language Processing*, pp.340-350(2021).
- [5] 柏木潔, 小町守, 松本裕治 : レビュー文書からの省略された属性の推定を含めた意見情報抽出, 言語処理学会第 19 回年次大会, pp.528-531(2013).
- [6] 楽天グループ株式会社 : 楽天市場データ, 国立情報学研究所情報学研究データリポジトリ(データセット), <https://doi.org/10.32130/idr.2.1>(2020).
- [7] MeCab, <https://taku910.github.io/mecab/>.

A 属性・評価値の出現回数

A.1 属性・評価値の出現回数上位 10 件

各データセットの属性・評価値の出現回数上位 10 件を表 8, 表 9 に示す。なお, 属性は文中に明記されている属性のみカウントした。

表 8 属性の出現回数上位 10 件

Tシャツ・カットソー	出現回数	セーター	出現回数
生地	394	生地	247
着心地	156	着心地	152
デザイン	110	肌触り	135
肌触り	81	デザイン	120
袖	63	形	102
丈	63	袖	99
素材	53	色	95
色	51	丈	77
形	49	ニット	66
サイズ	45	素材	62

表 9 評価値の出現回数上位 10 件

Tシャツ・カットソー	出現回数	セーター	出現回数
良い	403	良い	497
可愛い	252	可愛い	303
しっかり	141	奇麗	127
有る	125	暖かい	114
着やすい	108	有る	113
暖かい	102	着やすい	103
ゆったり	95	しっかり	85
薄い	86	ゆったり	65
綺麗	73	薄い	64
安い	67	柔らかい	59

なお, 全属性の合計出現回数は Tシャツ・カットソーが 3,125 回, セーターが 3,629 回である。また, 全評価値の合計出現回数は Tシャツ・カットソーが 4,023 回, セーターが 4,708 回である。

A.2 出現回数ごとの属性数・評価値数

各データセットの出現回数ごとの属性数・評価値数を表 10, 表 11 に示す。

表 10 出現回数ごとの属性数

Tシャツ・カットソー	属性数	セーター	属性数
100 回以上	3	100 回以上	6
50 回以上	5	50 回以上	6
100 回未満		100 回未満	
30 回以上	6	30 回以上	3
50 回未満		50 回未満	
10 回以上	37	10 回以上	49
30 回未満		30 回未満	
5 回以上 10 回未満	55	5 回以上 10 回未満	65
2 回以上 5 回未満	190	2 回以上 5 回未満	200
1 回	500	1 回	584

表 11 出現回数ごとの評価値数

Tシャツ・カットソー	評価値数	セーター	評価値数
100 回以上	6	100 回以上	6
50 回以上	6	50 回以上	6
100 回未満		100 回未満	
30 回以上	11	30 回以上	16
50 回未満		50 回未満	
10 回以上	39	10 回以上	47
30 回未満		30 回未満	
5 回以上 10 回未満	56	5 回以上 10 回未満	79
2 回以上 5 回未満	200	2 回以上 5 回未満	225
1 回	532	1 回	628