

レビュー文における的外れ文抽出の試み

屋比久 博文[†]琉球大学 工学部知能情報コース[†]當間 愛晃[‡]琉球大学 工学部知能情報コース[‡]

1 はじめに

私たちは生活の中で「何かを購入する」、「噂の映画を見に行く」など様々な行動を起こす以前にこれらの行動に対しての情報を収集することがある。その情報収集の手段として多く利用されているものが YouTube や Twitter などの SNS、食べログなどのレビューサイトである。しかしながらこれらの SNS やレビューサイトでは悪質なアンチによる動画や商品などと関係のないクレーム、動画を視聴していない人物や商品を利用していない人物などによる的外れな意見が存在する。こういったコメントやレビューは別の利用者にとって役に立たず、動画の投稿者や商品の提供者にとっても損失が生じる可能性がある。

そこで本研究では、このような的外れ文抽出を目的とし事前学習済みの BERT モデルをファインチューニングする手法を検討した。

2 関連研究

レビュー文を評価する関連研究として、山澤らが取り組んだ「Amazon レビュー文の有用性判別実験 [1]」が挙げられる。山澤らの研究ではレビュー文に対し有用/有用でないのラベル付与を手で行いこれらの判定を 2 値分類の機械学習手法であるサポートベクターマシン (SVM) を用いて行なった。しかしこのアノテーション方式における有用である/有用でないのラベルは付与

者の主観によって付与した結果にブレが生じることが問題として挙げられる。本研究では趣旨からそれたレビューに対する評価を的外れという言葉を用いることでラベル付与者の主観の介入をできるだけ避けた分類モデルの構築を目指す。しかしながら「的外れ」でも主観を除外することは難しかったため、後述の 4 ラベルによるアノテーションを採用した。

3 提案手法

本研究では事前学習済みの東北大 BERT モデル ('cl-tohoku/bert-base-japanese-whole-word-masking') をラベル付与を行ったレビューを用いてファインチューニングに取り組む。レビュー収集の対象としては、比較的多くのレビューが集まると考えられる YouTube のコメント欄を利用し、動画に対するコメントが的外れであるかどうかを判定することで評価を行った。

3.1 ラベル検討

YouTube のコメントに対するラベル付与アンケートを実施した。この結果、的外れにはラベル付与者が所持する動画外の専門的知識の有無によつて的外れの判断ができず、付与ラベルが変化することがわかった。このことからラベル付与に生じるブレ軽減のため、的外れの判断に動画外の専門的知識を必要とする場合に付与する“知識”ラベルを追加した。またコメントには短文で内容とは直接関係のないものが存在しこれらは解釈の仕方、つまり文章の補完によつて的外れかどうかの判断が変化する場合も存在した。このようなコメントに対応するため、“知識”ラベルと同様に“解釈”ラベルを追加した。これらの“的外れ”、“知識”、“解釈”ラベルは同時に起こる可能性がある。しかしながら知識や解釈の有無とそれに伴う

A trial of extracting out-of-focus sentences for review text

[†] Hirofumi Yabiku, Computer Science and Intelligent Systems Program, School of Engineering, Faculty of Engineering

[‡] Naruaki Toma, Computer Science and Intelligent Systems Program, School of Engineering, Faculty of Engineering

的外れかどうかについては回答者によって意見がばらつき、共通化することができなかった。このため今回は他のラベルの有無を問わず、ラベル毎に独立したアノテーションをすることとした。

3.2 ラベル付与手順

以下の手順でラベル付与を行った。

- 動画の title, 内容, 登場人物の確認を行いラベルを付与する。
- 動画の話題と関係のないコメントは的外れとしラベルを付与する。
- 的外れの判定に動画外の専門的知識が必要なものは“知識”ラベル, 解釈によって判定が変化するものには“解釈”ラベルを付与する。
- 動画に対するコメントや感想は的外れではないとしラベルを付与する。
- 全てのコメントに一度目を通しラベルを付与する。

4 実験

4.1 実験手順

実験手順を以下に示す。

1. YouTube 動画「理系レベル 1 と理系レベル 99 の違い [2]」より YouTube Data API を用いてコメント収集を行う。これにより 1159 件のコメントを収集した。
2. 取得したデータに対してアノテーションを行う。実装方法の都合上、アノテーションはラベルエンコーディングを行う必要がある。そこで本研究では的外れでない状態を表現するためラベルは、“的外れである可能性が高い”/“的外れでない可能性が高い”/“知識”/“解釈”の 4 つで行った。
3. 東北大 BERT モデル ('cl-tohoku/bert-base-japanese-whole-word-masking') を用いてファインチューニングを行う。学習はトレーニングデータ 6 割, バリデーションデータ 2 割, テストデータ 2 割とし, エポック数は 10 エポックの間バリデーションが改善されない場合に学習を終了する条件で行った。

4.2 実験結果

テストデータにおける 4 ラベルに対する抽出結果を表 1 に示す。

表 1 各ラベルに対する recal, precision, f1 score

	recal	precision	f1 score
的外れでない可能性が高い (1202)	1	0.919	0.957
的外れである可能性が高い (47)	0.625	0.833	0.714
知識 (71)	0.222	1	0.363
解釈 (19)	0	0	0

5 考察と今後の展望

表 1 の recall の値からの的外れである可能性が高いコメントと的外れの判断に知識が必要であるコメントは実際の正解ラベルを網羅できていないことが読み取れる。これは、的外れである可能性が高いコメントの判定に知識が必要となるコメントの特徴をうまく抽出できていないことが原因だと思われる。これに対する対策としては、多値分類ではなくマルチラベルとして“的外れである可能性が高い”と“知識”の 2 つのラベルを同時に考慮したモデルを構築することで改善できると考えている。また解釈のラベルは全データ 1159 件中 19 件と極端に少ないため特徴をうまく学習できておらずデータ拡張の必要があると考えている。

今後の展望としてコスト関数の実装, データの拡張, マルチラベル化をすることで分類の精度を向上させることを考えている。また本実験では一本の動画のみで学習を行ったため今後は複数の動画に対して学習させ汎用的な分類が可能になることを目指す。

参考文献

- [1] 山澤美由紀, 吉村宏樹, “Amazon レビュー文の有用性判別実験”, 情報処理学会研究報告自然言語処理 (NL), 53(2006-NL-173) 号, pp.15-20.
- [2] QuizKnock, 「理系レベル 1 と理系レベル 99 の違い」, YouTube, <https://www.youtube.com/shorts/OJ--u2fKTFU> (参照日 2023 年 1 月 9 日).