E-008

動画投稿サイトの動画からの特徴抽出によるユーザの意思決定支援に関する研究

Research for User Decision Support by Extracting Features from Videos on Video Posting Sites

堂前 拓生† 上田 芳弘† 坂本一磨†
Takumi Doumae Yoshihiro Ueda Kazuma Sakamoto

1. 緒言

近年，YouTube やニコニコ動画等の動画投稿サイトには，多くの動画が投稿され，様々な情報が発信されている．その中でも製品レビュー動画には，動画投稿者がある製品をどのように評価したのかという情報が多く含まれている．製品レビュー動画は，動画時間が長尺なものが多く，製品選択の際に複数の動画を視聴すると多くの時間を消費する傾向がある．そのため，製品レビュー動画から製品の特徴を抽出し，可視化することができれば，製品の購入を検討している視聴者にとって効率的な製品選択が可能となる．本研究では，製品レビュー動画の会話内容を文字起こしすることでその内容に対して，AI モデルを用いた自然言語処理や音響特徴量を用いた感情分析を行う．これにより，EC サイトのレビューのように製品選択の参考となる情報を提供することを目指す．なお，自然言語処理には，文章分類において優れたモデルである BERT（Bidirectional Encoder Representations from Transformers）を用いた．BERT の特徴としては，文章の分類問題，穴埋め問題等の複数タスクへの対応が，少ないデータによるファインチューニングで可能なことや，文章を双方向の Transformer によって学習することにより，文脈理解に優れた処理が可能であることが挙げられる．

2. 実験手順

本研究では以下のような手順で実験を行った．

(1) 製品レビュー動画に対して文字起こし
(2) 文字起こしされた文章を分割
(3) 分割した文章をカテゴリごとに分類

製品レビュー動画の文字起こしは，Reazonspeech，Whisper を使用している．Reazonspeech，Whisper で文字起こしした文章は，例えば，40 万円を 40 と万円で分割されるような不自然な部分で分割されている場合がある．そのため，文章ごとに分割する際，不自然な区切りにより意味が読み取れないことを防ぐ必要がある．その一方策として，BERT によって全ての文字の後を[MASK]し，そこに句読点が入るべきかを予測することで自然な句読点を挿入して，その読点で区切るという方法で文章分割を試行した．また，文章の分類には，BERT，BERT の改良モデルを使用する．分類カテゴリは，価格，性能，デザイン，耐久性，使いやすさ，その他の 6 つに分類している．このカテゴリについて適合率，再現率，F 値を算出し，性能を比較する実験を行う．

3. 実験結果

3.1 実験 1

実験 1 では，製品例としてカメラについての分類を行っている．使用した動画数は 15 本，センテンス数は各カテゴリ 40 センテンスで合計 240 センテンスである．BERT によって，文章分割したデータでは，動画数は同じでセンテンス数は，各カテゴリ 15 センテンス合計 90 センテンスである．文字起こしのモデルは Reazonspeech の v2 モデルを用いた．分類モデルは，BERT の中で文章に特化したと言われている SentenceBERT[1]を用いた．Reazonspeech を用いて，文字起こしをした文章をカテゴリごとに分類した結果を表 1，センテンスごとに分割した文章を分類した結果を表 2 に示す．なお，ロスを可能な限り下げるために Reazonspeech を用いた文字起こしの場合は学習率 10^(-6),学習回数 400 回とし，センテンスごとに分割した文章では，学習率 10^(-5)，学習回数 150 回とした．各実験結果から価格，耐久性に関しては再現率，適合率，F 値ともに高い値を示しているが，他のカテゴリでは，文字起こしの場合とセンテンスごとに分割した文章の場合の実験結果を比べると後者の方が高い値になる傾向が多いことがわかった．

表 1 Reazonspeech による分割の場合の分類精度

	再現率	適合率	F 値
価格	0.90	0.75	0.82
性能	0.50	0.47	0.48
デザイン	0.40	0.36	0.37
耐久性	0.70	0.88	0.75
使いやすさ	0.34	0.35	0.34
その他	0.28	0.35	0.30

表 2 BERT による分割の場合の分類精度

	再現率	適合率	F 値
価格	1.00	1.00	1.00
性能	0.52	0.73	0.59
デザイン	0.56	0.50	0.50
耐久性	0.64	0.95	0.76
使いやすさ	0.56	0.38	0.44
その他	0.50	0.57	0.53

3.2 実験 2

実験 2 では，実験 1 よりさらにデータを追加し，実験を行った．使用した動画数は 26 本，センテンスは，各カテゴリ 40 センテンス，合計 240 センテンスである．BERT，SentenceBERT，RoBERTa[2]，DeBERTa[3]という BERT の改良モデルを用いて実験を行った．BERT で検証した各カテゴリの精度を表 3 に示す．表 3 の実験結果より，この 4 つのモデルの F 値の平均は BERT が最も高い値になるという結果となった．

†公立小松大学 Komatsu University

3.3 実験 3

実験 3 では，文字起こしのモデルを Whisper の largev3 を用いて，データを増やすためにカメラだけでなく iPhone，イヤホン，家電，などの複数の製品のデータを用いて実験を行った．BERT のモデルの 1 つである BERT_largev2_を用い，学習率 10^(-6)，学習回数 50 回，1 文章中の最大単語数は 128 単語とした．Whisper の largev3 で文字起こしをしたデータでは，表 5 に示す結果となった．表 5 の結果より表 3 と比べて性能とその他のカテゴリでは，F 値が向上していたが，価格，デザイン，耐久性のカテゴリでは F 値が低下した．全体の F 値の平均としては，0.65 で同値であった．表 6 の結果より，単語数の最大を 128 単語から 256 単語にした場合は F 値が全体的に向上し，特に F 値の低かったデザインと性能のカテゴリは精度の向上が見られた．さらに正解を 1 つだけでなく上位 2 件のカテゴリを正解とした場合は正解率 88%という結果となっている．

4. 結論

各実験結果より分類精度が高い価格に関しては，「〇〇円」や「安い」などの具体的な価格を表す単語，また，耐久性に関しては，「電池持ち」や「温度に対して」など耐久性の特徴が分かりやすい単語が頻出していたため，再現率，適合率，F 値ともに高い値が示されたと考えられる．実験結果の分類精度が低い原因は，データ量が少ないこと，及び分割された 1 つの文章の中に複数のカテゴリの特徴語が含まれている文章が存在するため，意味に適合していない分類となっていることが考えられる．BERT，RoBERTa，SentenceBERT，DeBERTa というモデルを用いて実験を行ったが，BERT のモデルが最も今回の研究には，適したモデルであったと言える．その原因としては，BERT モデルの改良版として学習のデータの種類，学習の方法を変更しているが，それ以上に，より多くの言語データにおいて BERT のモデルが事前学習されているため，BERT のモデルの方が精度の高い結果が得られたと考えられる．表 5 と表 6 の比較において，表 6 で精度が向上しているのは最大の単語数が 128 単語から 256 単語にすることで 1 つの文章から得られる情報が増えたことで全体的に精度の向上が見られたと考えられる．特に，性能，デザインのカテゴリでは，新たに増えた情報の中に正しく判定するための情報が多かったと考えられる．また，文章内には 1 つのカテゴリの意味だけでなく複数の意味が含まれている場合があり，表 6 の実験結果で上位 2 件を正解とした場合の正解率は高い値となっているので，カテゴリの予測を 1 つでなく，上位 2 件とすることも検討すべきであると考えられる．

今後の課題としてはデータ量を増やすこと，並びに動画の文字起こしモデルの変更と，更なる高度な意味解析を用いた文章分割などにより分類精度の向上を目指す．今後の展望としては，分類後に動画のセリフの感情分析を行うことによって視聴者に製品の特徴を点数評価して可視化することで視聴者の意思決定支援に役立てていきたいと考えている．感情分析の方法としては，自然言語処理による分析だけでなく，音響特徴量を用いた分析によって感情分析を行うことも予定している．

表 3　データ追加後の BERT による分割の場合の分類精度

	再現率	適合率	F 値
価格	0.97	0.81	0.87
性能	0.37	0.59	0.45
デザイン	0.70	0.67	0.68
耐久性	0.70	0.80	0.73
使いやすさ	0.70	0.62	0.66
その他	0.70	0.53	0.53

表 4　分類精度（平均値）の比較

モデル	再現率	適合率	F 値
BERT	0.69	0.67	0.65
SentenceBERT	0.53	0.53	0.52
RoBERTa	0.45	0.47	0.50
DeBERTa	0.32	0.31	0.39

表 5　複数製品を対象とした場合の分類精度（1 文章中に最大 128 単語）

	再現率	適合率	F 値
価格	0.78	0.83	0.81
性能	0.61	0.5	0.55
デザイン	0.43	0.56	0.48
耐久性	0.60	0.75	0.67
使いやすさ	0.68	0.64	0.66
その他	0.78	0.65	0.71

表 6　複数製品を対象とした場合の分類精度（1 文章中に最大 256 単語）

	再現率	適合率	F 値	全体の正解率	上位 2 件正解のときの全体正解率
価格	0.88	0.81	0.84	0.71	0.88
性能	0.68	0.57	0.62		
デザイン	0.57	0.68	0.62		
耐久性	0.70	0.78	0.74		
使いやすさ	0.75	0.77	0.76		
その他	0.70	0.68	0.69		

参考文献

[1] Nils Reimers, Iryna Gurevych，Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks，arXiv:1908.10084，Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)，pp，3982–3992，(2019).

[2] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, Veselin Stoyanov，RoBERTa: A Robustly Optimized BERT Pretraining Approach，arXiv:1907.11692，(2019).

[3] Pengcheng He, Xiaodong Liu, Jianfeng Gao, Weizhu Chen，DeBERTa: Decoding-enhanced BERT with Disentangled Attention，arXiv:2006.03654，(2021).