

# 学 位 論 文 要 旨

学 位 論 文 題 名    Biterm Topic Modelを用いた商品レビューの関連性評価

(ふりがな)    にしはら りょうすけ  
学位申請者    西原 涼介

## ( 学位論文要旨 )

近年, Amazonや楽天市場などの大手ECサイトをはじめ, 数多くのECサイトが普及しその利用者数は急速に増加している. そして, 商品を購入する際にECサイトのレビューを参考にしている利用者の割合は約70%と言われていて, その中でもレビューの信頼性を重要視している人が多いことが明らかになっている. また, 多くの企業にとって, ECサイトのレビューからユーザーの嗜好や意見を分析し, マーケティングに活用することが重要な課題となっている. そのため, ECサイトのレビューの信頼性や有用性を評価する評判分析や口コミ分析, レビューを様々なトピックに分類する文書分類に関する研究が盛んに行われている. 例えば, 関連研究の項で詳しく紹介する「機械学習を用いた自然言語処理による商品レビューの評価」[1]では, Amazonの商品レビューを機械学習を用いて分析し, 参考になる順に並び替えるシステムの構築, 及びその評価に関する研究を行っている. また近年では, 従来のECサイトや商品のWebページだけではなく, YouTubeのような動画投稿サイトやX(旧Twitter)やInstagramなどのSNSを自社製品やサービスの宣伝の場として利用している企業が増加している. これに伴い, 商品を購入する際にSNSやYouTube上でその商品を宣伝している投稿を参考にしている人も増加している. そのため, SNSやYouTube上の広告に対するユーザーのコメントも, 他のユーザーが商品の購入を検討する際の重要な判断材料になり得ると考えられる. つまり, SNSやYouTube上での商品の宣伝に対するコメントは, ECサイトの商品レビューと同等の機能を持ち, その信頼性や有用性が重要視されるため, 同じく評判分析や文書分類の研究の対象になると考えられる. 例えば, SNSやYouTubeは商品レビューのページとは異なり, 誰でも気軽にコメントを投稿できたり, その投稿内容も自由という特性上, 商品やサービスに関係ないコメントが多数存在するという問題がある.

そこで, 本研究では分析対象をYouTube上で自社製品やサービスを宣伝している動画に対するユーザーのコメントとし, その動画で宣伝している商品やサービスに対しての関連性が高いコメントを抽出するシステムの作成, 及び作成したシステムの人手に対する精度の検証を目的としている. これにより, ECサイトの商品レビューと同様に他のユーザーが商品の購入を検討する際の重要な判断材料として扱うことが可能になると考えられる.

この目的の実現のため, 本研究ではトピックモデルの一種であるBiterm Topic Model(BTM)を用いたトピック抽出, 及び各トピックの生成確率上位の単語を利用した手法を提案する. BTMは

他のトピックモデルとは異なり、文書全体のバイターム(2単語の対)の共起性を利用してトピックを学習するため、YouTubeのコメントのように一文が比較的短い場合でも適切にトピックを推定することが可能であると考えられる。はじめに、商品やサービスの宣伝を行っているYouTube動画に対するコメントをYouTube Data APIを用いて取得し、クリーニング処理や分かち書きなどの前処理を施す。次に、前処理をしたコメント集合にBTMを適用し、潜在的なトピックの推定と各トピックにおける生成確率上位の単語を抽出する。そして、大規模言語モデルであるGPT-4を用いて、各トピックから抽出した単語を基に文章を自動的に生成する。ここで生成した文章はBTMによって推定したトピックに出現しやすい単語を用いているため、各トピックの代表的な文章であるという仮説を立てることができる。つまり、生成した文章は元のコメント集合に対して代表的であり、動画内容に関連している文章であるという仮説が立つ。その仮説を基に、トピックごとに生成した文章と元のコメント文との文章間の類似度を計算し並び替えることで、本研究の目的である動画に対して関連性が高い文章の抽出が実現できると考えた。また、本仮説の妥当性、及びシステムの精度を検証するために、人手でアノテーションしたデータセットとの比較分析を行った。

実際のYouTubeコメントを本手法に適用し実験を行なった結果、BTMによって商品に関連しているトピック及び単語を適切に抽出することができた。また、人手でアノテーションしたデータと本手法で類似度計算を行ったデータからConfusion Matrixを算出し、様々な指標で本手法の性能を評価した。その結果、Precision(適合率)が0.7程度の値を示したことから、本手法により商品との関連性が高いと予測したコメントのうち約7割が人手の評価に対して正しく予測できていることが分かった。この結果より、本研究で提案したSNSやYouTubeコメントに対する新たな分析手法の有効性を示すことができた。

(注) 2 0 0 0 字程度

学位論文題名が英文の場合は、和訳を( )に記入すること。