

商品紹介動画に対するコメントの関連性評価システムの作成

*西原 涼介 相馬 隆郎（東京都立大学大学院）

Creation of a System for Evaluating the Relevance of Comments to Product Introduction Videos

Ryosuke Nishihara, Takao Souma, (Graduate School of Tokyo Metropolitan University)

Abstract : EC サイト等で商品が宣伝・販売されているのと同様に、近年では企業が YouTube や SNS のアカウントを開設し、自社製品やサービスの宣伝を行なっている。その宣伝に対する消費者のコメントは、口コミサイトにおけるレビューと同等の機能を持ち、消費者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。しかし、SNS や YouTube では比較的気軽に投稿できるという特徴から商品に関係ないコメントが多数存在する。そこで本研究では、YouTube のコメントに対して対象商品との関連性を定量的に求める手法を提案する。具体的には全コメントに対して Biterm Topic Model を用いて各トピック上位の単語を抽出し、それらの単語と各コメントとの類似度を計算することで関連度を求める。また、人手でアノテーションした正解データと類似度上位の原文を比較し、作成したシステムの精度を検証した。

キーワード : 自然言語処理, 評判分析, 口コミ分析, バイタームトピックモデル

(Natural language processing, Reputation analysis, Review analysis, Biterm Topic Model,)

1. はじめに

近年、Amazon や楽天市場などの大手 EC サイト以外にも多くの EC サイトが普及し始め、その利用者也急増している。そして商品を購入する際に EC サイトのレビューを参考している利用者の割合は約 70%と言われている。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、企業にとっても 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を評価する評判分析や口コミ分析に関する研究が多く行われている。EC サイトのレビューの他にも、近年では YouTube のような動画投稿サイトや SNS での自社製品・サービスの宣伝を行う企業が増えてきている。そして、その宣伝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 サイトのレビューと同等の機能を持ち、評判分析の対象になると考えられる。しかし、SNS や YouTube はその特性上、誰でも気軽にコメントを投稿できるため、商品やサービスに関係ないコメントが多数存在する。そこで、本研究では Biterm Topic Model を用いた商品に関するトピック抽出によって、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から、その商品やサービスに対して関連性が高いコメントを抽出するシステムを作成した。

2. Biterm Topic Model (BTM)

コメント集合から商品との関連性が高い単語を抽出する手法として、文章から潜在的トピックを分析し抽出するトピックモデルが有効であると考えられる。従来のトピックモデルである Latent Dirichlet Allocation (LDA)[1]では、YouTube のコメントのように比較的短い文章の場合、データのスパース性が問題になり推定精度が低下してしまう。そのため、短い文章に対してもトピックを適切に推定可能なモデルとして、Biterm Topic Model (BTM)[2]が提案されている。本研究において、BTM は YouTube のコメントに含まれる 2 つの単語のペアをバイタームと定義し、バイタームごとに同一のトピックを仮定したモデルである。式 (1)、図 1 はそれぞれ BTM の確率モデルとグラフィカルモデルを示したものである。

$$P(B|\Theta, \Phi) = \prod_{j=1}^{N_B} \sum_{k=1}^K \theta_k \phi_{kx_{j1}} \phi_{kx_{j2}} \quad (1)$$

ここで、 K はトピック数、 N_B は総バイターム数、 Φ は単語分布、 Θ はコメント集合全体のトピック分布を表す。バイターム集合 B に含まれる各バイターム (2 つの単語 x_{j1}, x_{j2} のペア) は Θ に従って選択されたトピック z_j に対応する単語分布 ϕ_{z_j} に従って生成される。また、 α と β はそれぞれトピック分布と単語分布が従うディリクレ分布のハイパーパラメータである。BTM では崩壊型ギブスサンプリングによりパラメータ Θ, Φ を推定する。

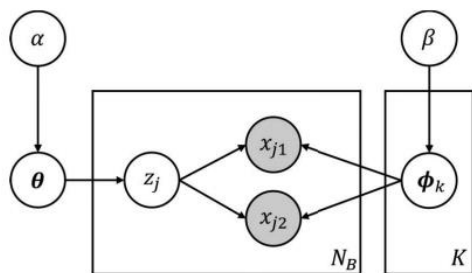


図 1：BTM のグラフィカルモデル

3. 提案手法

図 2 に本研究で提案するシステム図を示す．まず，YouTube の動画から取得したコメント集合に対して，BTM によって推定した K 個のトピックから生成確率上位 n 単語を抽出する．次に K 個のトピックごとに抽出した単語を基に文章を生成する．今回の実験では GPT-3.5 を用いて文章を生成した．生成した K 個の文章とコメント集合の一文一文との類似度を計算し，高い類似度を示したコメントから順に，商品との関連性が高いコメントであると評価する．文章間の類似度は，BERT の事前学習済みモデルによる文章のベクトル化とコサイン類似度によって計算する．また，商品との関連性の有無について人手で正解ラベルを付与したデータ M 件と類似度上位 M 件を比較し，その一致度を計算することで作成したシステムの精度を検証する．

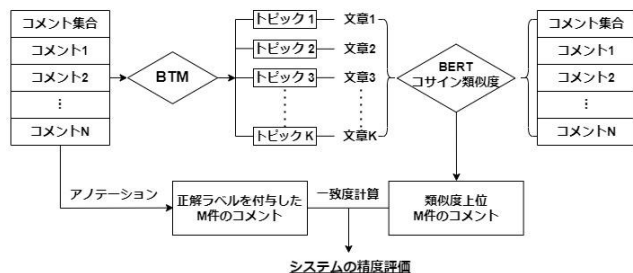


図 2：提案手法のシステム図

4. 実験

〈4.1〉 実験条件

本研究では，YouTuber のヒカキンが商品開発を行なった「みそきん（カップラーメン）」の発表動画[4]と，カズチャンネルが中部電力ミライズの提供で投稿した「こころモ（スマートリモコン）」の商品紹介動画[5]に対するコメントを実験の対象とした．コメントは YouTube Data API を利用し，それぞれ 1494 件，390 件取得した．BTM のハイパーパラメータは $\alpha = 0.1, \beta = 0.1$ ，トピック数は $K = 5$ ，抽出する単語数は $n = 10, 20, 30$ とした．

〈4.2〉 実験結果・考察

各動画に対して， $n = 10$ でトピック 1 の生成確率上位の単語を抽出した結果を表 1 に示す．表 1 より，それぞれの商品に関連性がある単語が抽出できているため，YouTube のコメントのような比較的短い文章においてもトピックを適切に推定できていることが分かる．

表 1：各動画の生成確率上位 10 単語

	みそきん	こころモ
1	味噌ラーメン	エアコン
2	味噌	電気代
3	ミサ	ガス
4	濃厚	カーテン
5	ヒカキン	温度
6	入っ	安く
7	好き	部屋
8	買う	使う
9	美味し	カズさん
10	笑顔	オール電化

次に，人手で正解ラベルを付与したデータと，抽出した単語から作成した文章との類似度を計算したデータを比較し，その一致度を算出した結果を表 2 に示す．

表 2：人手の正解データとの一致度

	$n = 10$	$n = 20$	$n = 30$
みそきん	0.65333	0.65777	0.65888
こころモ	0.78695	0.78260	0.76521

表 2 より，提案手法は約 65%~78% の精度で商品に関連性があるコメントを抽出できることが分かる．抽出する単語数による一致度の差は，当該コメントの長さ起因すると考えられる．また，「みそきん」に対するコメントの中で，正解ラベルを付与したが類似度が低く算出されたものとして，“ちぢれ麺がいいね”や“そのうち袋麺も出してくれ”といったコメントがある．“ちぢれ麺”，“袋麺”を含むコメントはどちらも 1 件のみであり，このように極端に出現回数が少ない単語を含む文は適切に抽出できていないと考えられる．

5. まとめ

本研究では，YouTube 上での商品紹介動画に対するコメントから，商品に関連するコメントを抽出するシステムを提案した．また，人手で作成したデータと比較し提案手法の精度を検証し，人手によるアノテーションなどの作業を必要としない手法で約 65%~78% の精度で抽出できるため，提案手法の有効性を示すことができた．今後の課題として，コメント総数が少ない動画に対する検討や，文章の長さを考慮した類似度計算方法の提案などが挙げられる．

参考文献

- [1] D. Blei, A. Ng and M. Jordan :“ Latent Dirichlet Allocation ”, Journal of Machine Learning Research, Vol.4, No.3, pp.148-159, 2019.
- [2] Xiaohui Yan, Jiafeng Guo, Yanyan Lan, Xueqi Cheng, “ A Biterm Topic Model for Short Texts”, WWW '13 Proceedings of the 22nd international conference on World Wide Web, pp.1445-1456(2013).
- [3] 西田 有輝, Tianxiang Yang, 山下 遥, 後藤 正幸 :「強調データの拡張学習による Biterm Topic Model の解釈性向上法に関する一考察」, 人工知能学会全国大会, 第 36 回, 3E3-GS-2-02, 2022.
- [4] <https://www.youtube.com/watch?v=SyTY7-ZTens>, (参照 : 2023-04-28)
- [5] <https://www.youtube.com/watch?v=Br5xQqKaMJI>, (参照 : 2023-06-08)