

PS8-2 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

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

*西原涼介 相馬隆郎(東京都立大学)
Ryosuke Nishihara, Takao Souma, (Tokyo Metropolitan University)

1. 研究背景・目的

インフルエンサー、芸能人への企業案件の増加
企業のYouTube, SNS利用数の増加

動画やSNSで自社製品やサービスの宣伝

コメントが口コミサイトのレビューと同等に
消費者の購入判断材料になる

欲しいものの情報はどこ入手？ ※男性 304 名、女性 333 名			
TV	48.0%	TV	52.3%
店舗	29.3%	店舗	26.1%
SNS	37.5%	SNS	59.5%
ウェブサイト/ 口コミサイト/ブログ	59.2%	ウェブサイト/ 口コミサイト/ブログ	57.1%
アプリ	22.7%	アプリ	22.8%
友人・家族	17.8%	友人・家族	24.3%
雑誌	10.2%	雑誌	10.8%
その他	5.3%	その他	2.1%

SNSやYouTubeは誰でも気軽に投稿できるという特性上、商品との関連性が低いコメントも当然存在する

図1: 情報収集方法に関するアンケート結果
ECのミカタ, 「ECとSNSに関する調査」,
<https://ecnomikata.com/column/26944/>より引用

研究目的

YouTube上の商品紹介動画に対するコメントと商品との関連性を評価し、
消費者の購入判断材料になるコメントを抽出するシステムを作成する

2. Biterm Topic Model

従来モデル(LDA)

- 文書レベルでの単語の出現頻度などをもとにトピックを学習する
- 長い文章であれば多くの単語が含まれるので比較的学習しやすい
- 短い文書ではデータのスパース性が問題になる

Biterm Topic Model

- 文書全体のバイタームの共起性を利用してトピックを学習する
- 文書レベルでのスパース性の問題を解決し、短い文書でも適切にトピックを推定する

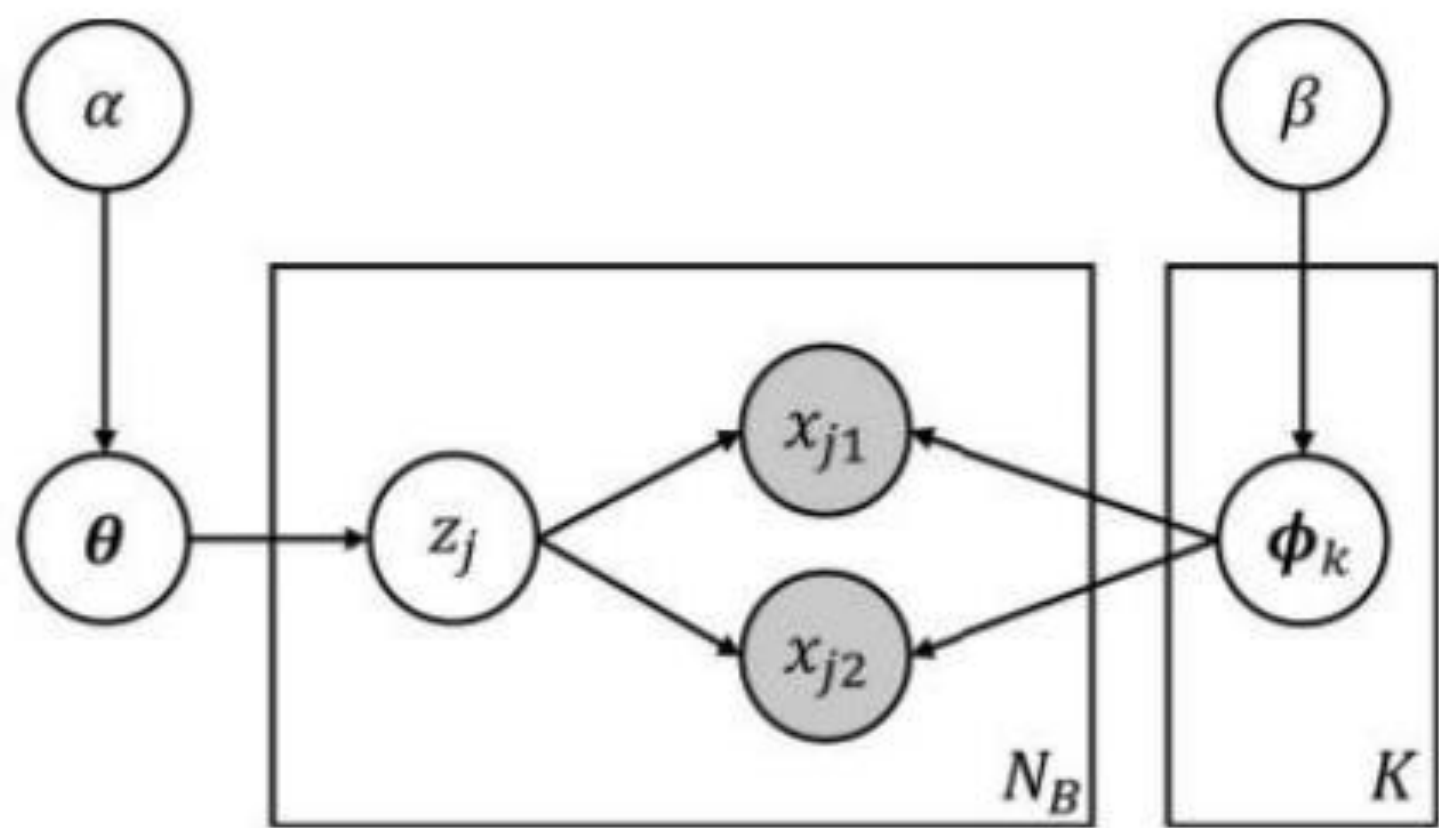


図2: BTMのグラフィカルモデル, 3.より引用

K : トピック数, N_B : 総バイターム数, ϕ : 単語分布, θ : トピック分布, z : トピック, x : 単語

単語の共起(バイターム)を利用することで、
トピックの意味を学習する質を向上させる

文書全体で集約されたグローバルな共起を
利用することで、文書レベルでのスパース
性を解決する

式(1): BTMの確率モデル

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

3. 関連性評価システムの提案手法

- YouTube Data APIで対象動画のコメントを取得
- コメントの前処理
- 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
- K個のトピックごとに、抽出した単語をもとに文章を生成(GPT-3.5)
- 生成したK個の文章とコメント全文との類似度を計算(BERT)
- 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と評価
- 元コメントに対して、人手で商品との関連性の有無の正解ラベルを付与
- 正解ラベルを付与したM件のコメントと、類似度上位M件のコメントを比較し
提案手法の精度を検証

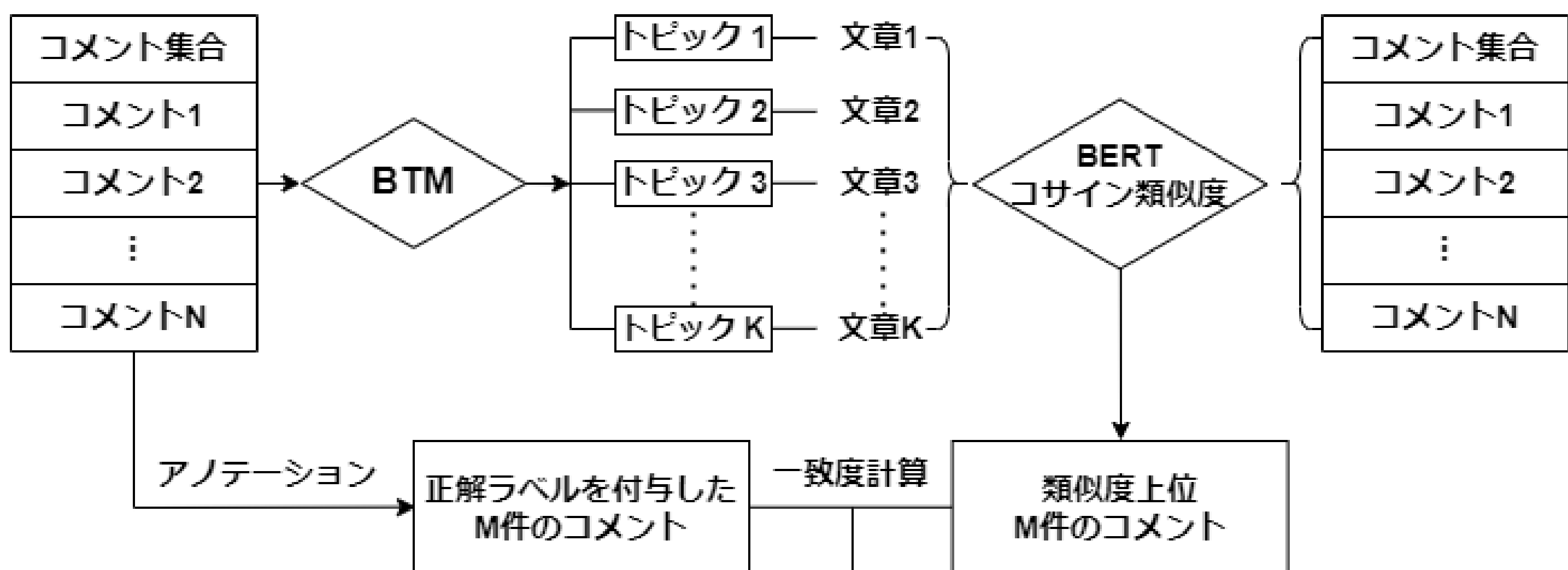


図3: 提案手法のシステム図

システムの精度評価

参考文献

- D. Blei, A. Ng and M. Jordan: "Latent Dirichlet Allocation", Journal of Machine Learning Research, Vol.4, No.3, pp.148-159, 2019.
- 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).
- 西田 有輝, Tianxiang Yang, 山下 遼, 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. 実コメントを用いた実験・結果

実験データ

- HikakinTV(みそきん)**
自身で商品開発を行ったみそきん
(カップラーメン)の発表動画
コメント取得日: 2023/04/28, 取得数: 1494件
- カズチャンネル(ここりモ)**
中部電力ミライズの提供でここりモ
(スマートリモコン)の商品紹介動画
コメント取得日: 2023/06/08, 取得数: 390件

前処理

- 絵文字、顔文字、記号、URL除去
- 分かち書き(MeCab)
辞書: mecab-ipadic-NEologd
- ストップワード除去
指示語、助詞、助動詞、形容詞など

BTMによるトピック抽出

パラメータ

$\alpha = 0.1, \beta = 0.1$, トピック数 $K = 5$
取得単語数 $n = 10, 20, 30$

みそきん: カップラーメンの味
や商品開発の際に参考にした店
舗に関する単語、視聴者の商品
に対する感情などの単語が取得
できている

ここりモ: 商品に対応している
家電に関する単語や、使用した
際の電気代やガス代などに関する
単語を取得できている

自動生成した文章例

みそきん: **ヒカキン**さんの**笑顔**が、**味噌ラーメン**を食べるときの私の楽しみです。 **濃厚**な**味噌**の風味が、心温まる**ミサ**と共に**入っ**た一杯は、本当に**美味**しいんですよ。いつも**ヒカキン**さんの動画でラーメンの話をする姿を見ると、つつい食べたくなくなってしまいます。

ここりモ: **カズさん**は**部屋**の**温度**を**安く**保つために、**エアコン**の使用を控え、代わりに**カーテン**を使っています。さらに、**オール電化**のシステムを導入しているため、**電気代**も**ガス**代も節約できています。

文章間の類似度

商品との関連性が高いコメントの類似度が高くなっているが、
類似度が高くても商品との関連性が低いものもある

類似度	みそきん 元コメント
0.94179	味噌ラーメン好きとしてとても気になる絶対買う味噌派閥にも嗜好が違う事があるからhikakinさんがどんな味が好きなのか楽しみ
0.92595	白味噌も入ったベースのニンニクのパンチあるラーメンの味の意味がわかりましたニューミサ僕も大ファンですコレは試してみたい
0.91985	再生回数見て思ったけど今でもみんなの関心を惹き付けてる事が凄いやんとみんなhikakinの事気になってるんだね
類似度	ここりモ 元コメント
0.91363	家にいる時間が増えてきている今とても便利な機能ですねスマホで出来るのがほんとに素晴らしい一人暮らしの方でも使いやすくて良いですね
0.91185	夏場は電気代高くなりますビックリします今の世の中家で過ごす事が多いので特にクーラー付けっぱなしですここりモは便利そうなので買ってみようと思いますカズさん良いものを紹介してくれてありがとうございます
0.89938	カズさんいつも動画作りお疲れ様ですカズさんにお聞きしたいのですが0円前後でふわふわのかき氷が作れるかき氷機を教えてください出来れば自動のがいいです

人手によるアノテーションとの比較

アノテーション基準

- 商品に関わるコメント
- 商品に対する視聴者の感情・意見
- 動画内容に関わるコメント

表2: アノテーション結果

	関連性あり	関連性なし
みそきん	900件	594件
ここりモ	230件	160件

各動画において“関連性あり”の正解ラベルを付けたコメントM件と、類似度上位M件の一致率を計算することで、提案手法の精度を検証する

表3: 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
ここりモ	0.78695	0.78260	0.76521

人手の評価に対して、**提案手法は65%~78%の精度**で商品との関連性が高いコメントを抽出できる

5. まとめ・今後の展望

BTMによるトピックと単語抽出
YouTubeのコメントのように比較的短い文章でもトピックを適切に抽出できる

人手による正解データとの一致率
提案手法は人手の評価に対して65%~78%程度の精度を示した

生成した文章との類似度計算
極端に出現回数が少ない単語を含む文は類似度が低くなる
関連性があるコメントでも、文章が極端に短いと類似度が低くなる

改善案
今後の検討
トピック数や抽出単語数による
文章を自動生成する方法を新たに考案
文章の長さを考慮した類似度計算法の考案
多くの動画に対して実験をし、提案手法の精度を検証