

顧客レビューによる購入目的別の品質要素比較分析モデル

Comparative analysis model of quality factors
by purpose of purchase based on customer reviewsトウ シセン[†] 山極 綾子[†] 邵 騰飛[†] 後藤 正幸[†]
Shisen Tou Ayako Yamagiwa Tengfei Shao Masayuki Goto

1. 研究背景・目的

近年、電子商取引サイトの普及により、消費者によるレビューが購買意思決定や商品改善において重要な情報源として注目されている。レビューには、評価スコアだけでなく、消費者の体験や期待、満足度に関する多くの定性情報が含まれている。これらを分析することで、企業は顧客の潜在的なニーズを把握し、よりの確な品質改善やマーケティング戦略の立案が可能となる。

こうしたレビュー分析に関する研究では、レビュー文に含まれる特徴語を抽出し、機械学習によって評価スコアとの関連を分析する手法が数多く提案されてきた。一方で、これらの多くは全レビューをまとめて扱っており、消費者の「購入目的」という視点に基づく分析は十分に行われていない。例えば、贈答用か否かによって、評価基準や重視する要素が異なる可能性が高く、それに伴ってレビューの内容やスコアへの影響要因も変化すると考えられる。たとえば、贈答用では包装の丁寧さや見た目の印象が重視される一方、自分用ではコストパフォーマンスや実用性が評価されやすい。しかし、こうした購入目的の違いがレビューにおいてどのように反映されるのか、また、各話題が評価スコアに与える影響が目的によってどのように変化するかについて、定量的に分析した研究は少ない。また、既存研究の多くはレビュー内のキーワードを単語レベルで抽出・分析するアプローチを取っており、文脈全体を踏まえた話題(トピック)単位での分析や、影響要因を比較可能な形で抽出する手法は限定的である。

そこで本研究では、Amazon における“Toys & Games”カテゴリのレビューを対象とし、購入目的を「自分用」と「贈答用」に分け、それぞれのレビュー群に含まれるトピックと評価スコアとの関係を分析することで、目的によって異なる評価構造の把握を目指す。具体的には、BERTopic を用いてレビューからトピックを抽出し、各レビューが各トピックに連続する度合いを確率分布として定量化する。その後、目的別に「高評価 vs 低評価」の2値分類タスクを設定し、トピック分布を説明変数とするロジスティック回帰モデルを構築する。さらに、得られた各トピックの回帰係数の差異を比較することで、各購入目的において満足度を押し上げる「魅力的品質」や、欠如によって満足度が下がる「一元的品質」といった品質要素の特性を明らかにする。これにより、同一商品であっても目的に応じて重視される品質要素が異なることを明らかにし、今後の顧客対応や商品開発に活用可能な知見を提供することが期待される。

2. 関連研究

近年、レビュー解析において、文書埋め込みを利用した手法が注目されている。Grootendorstら[1]は、Transformer による高品質な文書ベクトル生成と c-TF-IDF によるトピッ

ク語抽出の組み合わせを用いた BERTopic を提案し、複数ベンチマークで LDA や NMF よりも優れた性能が得られることを示した。さらに、Krishna ら[2]は顧客レビューに LDA・NMF・BERTopic を適用した結果を比較し、BERTopic が最も意味的に一貫したトピックを抽出できたと報告している。また Thom Lechaude ら[3]は Amazon レビュー(“beauty”カテゴリ)に対して BERTopic を適用し、レビュー内容の属性把握やユーザー意図の抽出に成功した。しかし、これらの研究はレビュー全体を統一的に処理しており、「購入目的」によってモデルを層別し、トピックと評価スコアの関連性を目的別に比較するという手法には至っていないという点で、本研究とは異なる。

2.1 BERTopic

BERTopic は、BERT によって得られる文書埋め込みを基盤とし、トピックモデリングを行う手法である。従来の LDA のように確率的な生成過程を仮定するのではなく、Transformer モデルによって高次元ベクトル化された文書表現をクラスタリングすることで、意味的に類似した文書群を導出する点に特徴がある。

BERTopic の基本的な流れは以下の通りである。まず、文書データに対して事前学習済みの BERT モデルなどを用いて文書ベクトルを取得する。このベクトルは各文書の意味的な特徴を反映しており、従来の BoW (Bag-of-Words) や TF-IDF よりも文脈的な意味を捉えることが可能である。次に、得られたベクトルを用いて次元削減(UMAP: Uniform Manifold Approximation and Projection)を行い、文書間の距離構造を保ちながら低次元空間に射影する。その後、HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise) によってクラスタリングを行い、トピックを定義する。各トピックに対応する代表語は、クラスターに属する文書から TF-IDF に類似した重み付け(class-based TF-IDF, c-TF-IDF)によって抽出される。

これにより、文書集合に対して意味的に整合性のあるトピックを自動的に抽出することが可能となる。また、各文書におけるトピックの分布(所属確率)も transform 関数によって得られるため、本研究のように得られた分布を特徴量とした回帰分析等への応用も容易である。

図 1 に BERTopic の概要を示す。

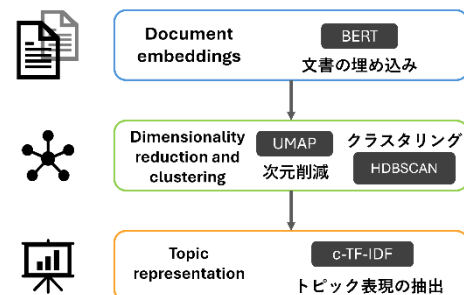


図 1 BERTopic モデルの概要図

2.2 ロジスティック回帰モデル

ロジスティック回帰モデルは、目的変数が2値（例：高評価／低評価）の場合に、説明変数との関係をモデル化するための手法である。本モデルでは、事象が発生する確率をロジスティック関数によって表現し、線形結合された説明変数に基づいて発生確率を推定する。

ロジスティック回帰のモデル式は以下のように表される：

$$\log\left(\frac{P(Y=1)}{1-P(Y=1)}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p \quad (1)$$

ここで、 Y は目的変数（例：高評価であれば1，低評価であれば0）， x_1, x_2, \dots, x_p は説明変数， β_0 は切片， $\beta_1, \beta_2, \dots, \beta_p$ は各説明変数に対応する回帰係数を表す。

3. 提案手法

本研究では、レビュー文に含まれるトピックと評価値との関係に着目し、購入目的による評価構造の差異を明らかにする分析手法を提案する。提案手法は以下の3つのステップから構成される：

Step1では、すべてのレビューに対してBERTopicを適用してトピック抽出を行う。Sentence-BERTによる文書ベクトル化、UMAPによる次元削減、HDBSCANによるクラスタリング、そしてc-TF-IDFによるトピック表現抽出を通じて、各レビューが複数のトピックに属する確率（トピック分布）を算出する。

Step2では、レビューに記載された購入目的（「自分用」「贈答用」）に基づきレビューを分類し、それぞれに対してロジスティック回帰モデルを構築する。説明変数には各レビューのトピック出現確率を用い、目的変数には評価スコアの高低をもとにした二値分類を設定する。

具体的には、以下の3種類のスコアスケールを適用する：

- ✓ 高評価5 vs 低評価1~4（最も厳しい基準）
- ✓ 高評価4~5 vs 低評価1~3（中間基準）
- ✓ 高評価3~5 vs 低評価1~2（最も緩やかな基準）

この3つのスケールを「自分用」「贈答用」のそれぞれに適用する。

Step3では、各モデルにおけるトピックごとの回帰係数の変化パターンを比較し、それがどのような品質要素に該当するかを考察する。例えば、あるトピックが厳しいスケール（5 vs 1~4）で大きく正の係数を示す一方で、スケールが緩やかになるにつれて係数が小さくなる場合、そのトピックは「魅力的品質」を表す要因であると解釈できる。一方で、すべてのスケールにおいて安定して正の係数を示す場合は「一元的品質」を表す要因に、負の係数が一貫して現れる場合は「当たり前品質」を表す要因に分類される。

このようにして、トピックの評価影響構造をスケール別・目的別に定量的に比較することで、各トピックの品質特性を明らかにし、それを狩野モデルの「魅力的品質」、「一元的品質」、「当たり前品質」に分類することを目指す。

提案手法の概要を図2に示す。

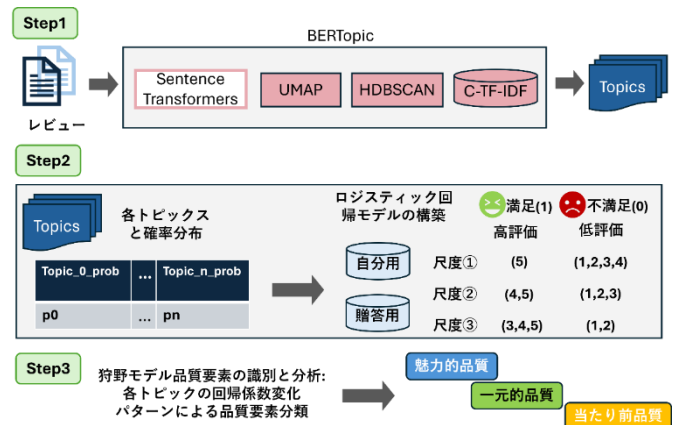


図2 提案手法の概要図

4. 実データ分析

4.1 データセット

本研究では、提案手法の有効性を検証するため、アマゾンのレビューデータセット[4]の“Toys & Games”カテゴリに属するレビューを対象として分析を行った。元データからランダムに10,000件のレビューを抽出し、レビュー本文の文字数が50~250字の範囲に収まるもののみを抽出した。その結果、最終的に7,070件のレビューが分析対象となった。本データセットは、以下の公開サイトより取得したものである（<https://amazon-reviews-2023.github.io/>）。

なお、本データセットには購入目的の情報が含まれていなかったため、本研究ではレビュー本文を観察し、「自分以外の人物（例：daughter, son）」への言及があるものを「贈答用」、それ以外を「自分用」として分類可能と判断した。この基準に基づき、大規模言語モデル（GPT）により一括分類を行い、一部を確認した上で「目的」列を補完し、分析を実施した。

4.2 実験設定

実験では、軽量かつ多言語対応のSentence-BERTモデル「all-MiniLM-L6-v2」を用いて各レビューをベクトル化した。出力された埋め込みベクトル（384次元）をあらかじめBERTopicの入力とした。

トピックモデリングには、UMAPとHDBSCANによる次元削減およびクラスタリングを組み合わせたBERTopicを用いた。UMAPモデルには $n_neighbors=5$ 、 $n_components=5$ 、 $min_dist=0.0$ など、局所構造を強調するパラメータを設定した。クラスタリングにはHDBSCANを用い、 $min_cluster_size=50$ 、 $min_samples=5$ とした。

また、テキストの高頻度語抽出には $max_features=3000$ 、 $min_df=3$ 、 $max_df=0.95$ をパラメータとするCountVectorizerを用い、英語のストップワードおよびドメイン特有の冗長語（例：“good”, “used”, “play”など）を除外した。

BERTopicモデルを適用し、各レビューに対するトピックの確率分布を出力させた。この分布を特徴量とし、ロジスティック回帰モデルにより評価スコアの二値分類を実施し

た。モデルの解釈には各トピックの回帰係数を用い、それぞれの品質要素が消費者の評価に与える影響を分析した。

4.3 分析結果

4.3.1 トピックの抽出と統合 (Step1)

まず、レビュー本文に含まれるトピックを抽出するため、BERTopic によるクラスタリングを実施した。全体で 46 個のトピックが得られたが、抽出されたトピックの中には、意味的に類似しているものや、トピック内のキーワードが重複しているものが複数見られた。トピック数が多すぎる場合後続の解釈や係数の比較が煩雑になるため、本研究では代表的なトピックに統合する処理を加えた。

トピックの統合には、BERTopic が出力する階層構造図 (図 3) を用いた。階層構造図では、類似度の高いトピックが色分けされた枝で接続されており、意味的に近いトピック同士を視覚的に判別できるようになっている。そこで、同じ色の枝で繋がったトピック群を一つのまとまりとして扱い、8 つの代表トピックに統合した。

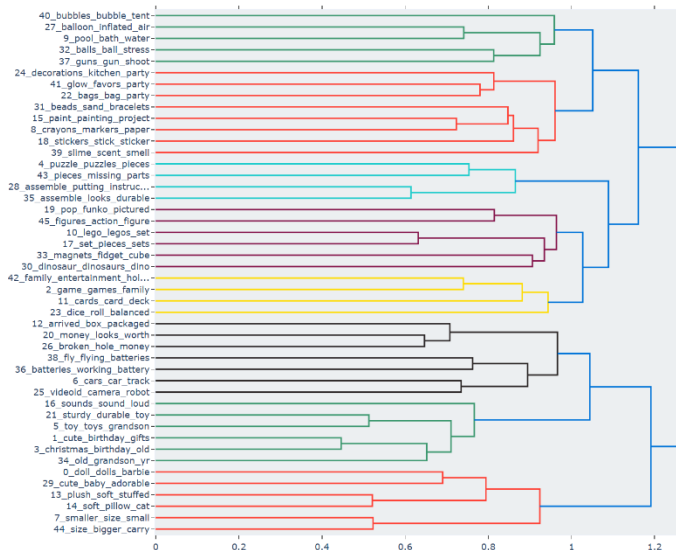


図 3 階層的クラスタ分析

統合後の各トピックについては、トピックごとに出現頻度の高いキーワードを抽出し、その特徴を簡潔に把握できるように可視化を行った (図 4)。例えば、トピック 0 では「toy」「doll」「grandson」「christmas」などの単語が多く出現しており、プレゼントや子どもへの贈り物に関連するレビューが中心となっていた。一方で、トピック 1 では「cards」「stickers」「perfect」「size」「color」などの語が確認され、小型商品の見た目や装飾に関する内容が主であった。

他のトピックとしては、「water」「balls」「bubbles」など、水遊びに使われる玩具に関する内容、「batteries」「waste」「return」など、電池の持ちや返品対応に関連する語が頻出していった。これらの結果から、各トピックは、レビューにおける具体的な使用経験や評価対象の側面に基づいて構成されていることが確認できる。このように、本研究では、Step2 のロジスティック回帰分析においてこれらのトピックを活用することを念頭に、各トピックの意味について人手による解釈を与えた。

なお、各レビューには複数のトピックが同時に含まれている場合があるため、トピックの出現有無ではなく、各レビューに対するトピックの確率分布を算出し、それを後続のモデル構築における説明変数として用いることとした。

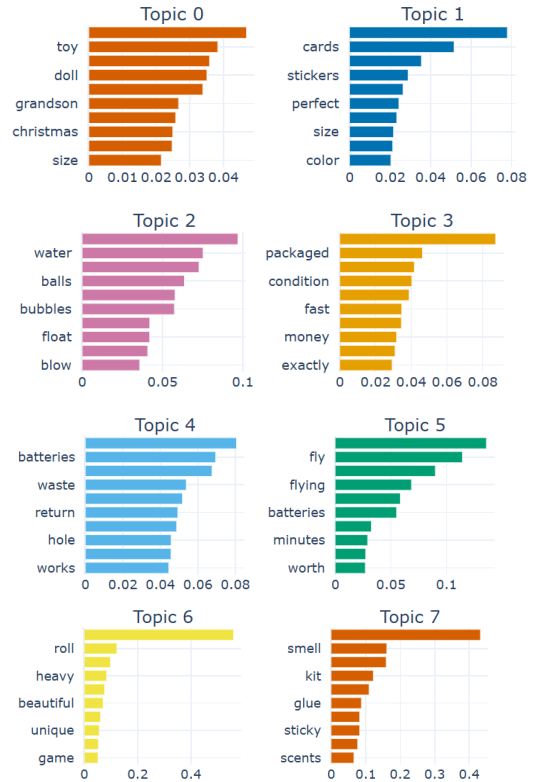


図 4 各トピックにおける代表キーワードと重要度

4.3.2 ロジスティック回帰モデルの構築 (Step2)

Step1 で得られたトピックの確率分布をもとに、ロジスティック回帰モデルの構築を行った。各レビューは、トピックに対する確率を特徴量として持ち、この確率分布を説明変数として用いた。

目的変数には、レビューに付与されている 1~5 の評価スコアに基づき、以下の 3 種類の二値分類ラベルを作成した。

- ✓ ラベル①: スコア 5 を 1, それ以外 (0~4) を 0
- ✓ ラベル②: スコア 4, 5 を 1, 0~3 を 0
- ✓ ラベル③: スコア 3~5 を 1, 0~2 を 0

これにより、各レビューは「目的」×「評価スケール」の組み合わせによって分類され、合計 6 パターンの回帰モデルが構築された。

4.3.3 係数に基づくトピックの分類 (Step3)

最後に、Step2 で得られた 6 つのロジスティック回帰モデルの出力を比較し、各トピックの係数の傾向に基づいて、トピックをタイプ別に分類した。具体的には、以下の基準により分類を行った：

- ① すべての評価基準と目的において正の係数を示し、満足度に一貫して寄与しているトピック (=一元品質)
- ② 評価基準が厳しい条件 (5 vs 1~4) で係数が大きく、基準が緩やかになるにつれて係数が小さくなるトピック (=魅力品質)
- ③ 主に負の係数を示し、存在しないことによって不満につながるトピック (=当たり前品質)

このようにして、8つのトピックについて、それぞれが「どのような評価条件で満足度に影響を与えるか」に応じて、一元的品質・魅力的品質・当たり前品質の3つに分類した。

表 1 自分用におけるトピック別の係数変化

Topic	自分用 _5vs1~4	自分用 _4,5vs1~3	自分用 _3,4,5vs1~2
Topic0	0.7259	1.0321	1.2689
Topic1	0.5910	0.6583	0.6169
Topic2	-0.3957	-0.5551	-0.5305
Topic3	0.3704	0.0855	0.1313
Topic4	-1.3837	-1.6388	-1.8747
Topic5	-0.3178	-0.5830	-0.4070
Topic6	0.7036	1.0605	1.0715
Topic7	-0.3561	-0.0345	-0.1718

表 2 贈答用におけるトピック別の係数変化

Topic	贈答用 _5vs1~4	贈答用 _4,5vs1~3	贈答用 _3,4,5vs1~2
Topic0	0.7445	0.8598	0.5067
Topic1	0.4276	0.2947	0.3152
Topic2	0.0235	-0.0406	-0.1461
Topic3	0.0352	-0.0211	-0.1105
Topic4	-1.7488	-2.2531	-2.9150
Topic5	-0.1067	0.1794	-0.2424
Topic6	0.1665	-0.0165	2.2721
Topic7	0.2628	0.9217	0.5174

自分用・贈答用それぞれのロジスティック回帰結果に基づき、トピックを狩野モデルの三要素に分類した。Topic 0 (玩具・プレゼント関連) と Topic 1 (カード・シール・外観) は、自分用・贈答用の双方において安定して正の係数を示し、一元品質として基本的な満足感を支えていることが確認された。Topic 3 (梱包状態・配送スピード) は、自分用においてのみ高評価時に強く寄与し、魅力品質として評価される一方で、贈答用では係数が低下していた。贈答用では、Topic 6 (美しさ・ユニークさ) や Topic 7 (匂い・手作り感) が魅力品質として機能しており、印象的・感性的な価値が重視されていた。一方、Topic 4 (バッテリー不良・返品・動作不良)、Topic 2 (水遊び・浮遊・シャボン玉)、および Topic 5 (飛行玩具・稼働時間) などは、全体的に負の係数を示し、当たり前品質として満足の前提条件となっていた。

自分用と贈答用を比較すると、一元品質 (Topic 0, 1) には共通点が見られ、どちらの目的でも製品の基本性能や外観の整いが満足度に安定して貢献していた。一方で、魅力品質の要素には明確な違いがあり、自分用では体験に関連す

るトピック (Topic 3) が高く評価されるのに対し、贈答用では視覚的・感性的な要素 (Topic 6, 7) が強く影響していた。また、トラブルや不良に関するトピック (Topic 2, 4, 5) は、自分用・贈答用を問わず満足度を著しく低下させており、どの目的でも避けるべき当たり前品質と位置づけられる。

自分用と贈答用を比較すると、一元品質に関しては共通しており、基本機能や外観の整いといった要素が満足度の土台を構成していた。対して、魅力品質の内容には明確な違いが見られ、自分用では使用体験に関する話題 (梱包や配送など)、贈答用では見た目や特別感を演出する要素がより強く評価される傾向があった。また、当たり前品質に該当するトピックはどちらの目的でも評価を大きく下げることが示され、満足以前に回避されるべき重要な要因であるといえる。

5. 考察

本研究では、レビュー文に含まれるトピックと評価値との関係性に着目し、BERTopic による話題抽出とロジスティック回帰モデルを組み合わせることで、各話題が満足度に与える影響を定量化する手法を構築した。さらに、自分用と贈答用という購入目的の違いを考慮してモデルを分割することで、それぞれの目的において評価されやすい品質要素の傾向を明らかにすることができた。

係数の変化パターンに基づき、トピックを狩野モデルの枠組みに沿って分類した結果、自分用では梱包や配送など体験的な話題が魅力的品質として評価され、贈答用では見た目の美しさや香りといった感性的な話題がより強く評価に寄与していることが示された。一方で、一部のトピックにおいては、目的や評価基準によって係数の傾向が異なるなど、単純に分類できない事例も見られ、今後はより文脈的な意味の分離やトピック数の調整が必要となる可能性もある。

6. まとめと今後の課題

本研究では、BERTopic によるトピック抽出とロジスティック回帰モデルを組み合わせることで、レビュー文に含まれる各トピックが評価スコアに与える影響を定量化する手法を構築した。さらに、提案手法を実データに適用し、「自分用」「贈答用」という購入目的の違いに着目した上で、各トピックの係数を比較し、狩野モデルに基づいて品質要素の特性を分類した。

一方で、トピック内部に複数の語義的特徴が混在している場合や、係数の変化傾向が明確でない場合には、品質要素としての分類が曖昧になるケースも見られた。今後の課題としては、トピックの語彙的分布やレビュー文中での使われ方をさらに精緻に把握することで、より安定した品質要素の抽出が期待される。

参考文献

- [1] Grootendorst, M., "BERTopic: Neural topic modeling with a class-based TF-IDF procedure", *arXiv preprint arXiv:2203.05794*, (2022).
- [2] Krishnan, A. Baby & Kennedyraj, "Exploring the Power of Topic Modeling Techniques in Analyzing Customer Reviews: A Comparative Analysis", *arXiv preprint arXiv:2308.11520*, (2023).
- [3] Léchau-dé, R. M. P., "Decoding Customer Feedback: Employing Topic Modelling (BERTopic) to Understand Product Return Drivers from Online Reviews", *M.Sc. thesis, Politecnico di Milano*, (2024).
- [4] Amazon Product Review Dataset 2023, <https://amazon-reviews-2023.github.io/>, (accessed 2024-06-01).