

顧客レビューを用いた電子製品の評価要因分析モデルに関する一考察

An Analysis of Evaluation Factors for Electronic Products Based on Customer Reviews

呂 文博[†] 阪井 優太[†] 邵 騰飛[†] 後藤 正幸[†]
 Wenbo Lyu Yuta Sakai Tengfei Shao Masayuki Goto

[†]早稲田大学大学院 創造理工学研究科 経営システム工学専攻

1. 研究背景と目的

近年、EC サイトにおける顧客レビューは、消費者の購買行動や商品の評価に大きな影響を与えている[1]. 特に星の数で満足度を付与した星評価は、簡易かつ直感的な評価指標として広く活用されており、EC サイト内の検索順位や推薦システム、売上にも密接に関連している[2]. 企業にとっては星評価の向上が重要なマーケティング施策の一つであり、特に星評価 3 から 4 への評価向上は売上改善に直結することが示されている[3]. 一方で、星評価 3 が付与されたレビューは内容の幅が広く、肯定的意見から批判的意見まで多様な感情が含まれている。

そのため、星評価 3 のレビュー内部に潜在する「ポジティブ評価」や「ネガティブ評価」の差異を精緻に把握することは、推薦システムの精度向上や商品改善にとって有用である。しかし、従来のレビュー分析では、ポジティブ/ネガティブ分類や星評価の平均値に基づく手法が主流であり、レビュー内部の細かな感情差異、特に星評価 3 帯に潜む「隠れポジティブ」「隠れネガティブ」に対するアプローチが行われてこなかった。

本研究では、特にデザイン・機能性・耐久性などの多様な要因がレビューに存在する電子製品領域に着目し、内容と星評価との関係性を分析する。手法を提案し、Amazon レビューの実データを分析した結果を示す。

2. 準備

2.1 関連研究

近年、商品レビューの感情分析には BERT や RoBERTa などの言語モデルが広く利用されている。Zhang ら (2021) は複数の評価基準を考慮した Learning to Rank 手法により、推薦システムのランキング精度向上を図っている[4]. また、Guo ら (2020) は RoBERTa を用いて Amazon レビューから細粒度な感情スコアを抽出し、消費者行動分析に応用した[5]. 一方、レビューの中立クラス分類は依然として課題であり、Taywade ら (2023) もその困難さを指摘している[6].

しかし、既存研究の多くはレビュー全体の単純なカテゴリ分類に留まっており、星評価 3 帯のレビューの感情差異を細粒度で可視化・分析し、実サービスに応用する試みは限られている。

本研究では、RoBERTa 感情スコア、SBERT 文ベクトル、BERTopic トピック情報を組み合わせ、LLM による弱教師ラベルを活用した星評価 3 帯のレビューに対する細粒度感情分析アプローチを提案する。

2.2 Sentence-BERT (SBERT)

SBERT は、BERT を基盤とした文レベル意味表現モデルであり、文間類似度を高精度に捉えることが可能である[7]. 本研究では、レビュー文を 768 次元の文ベクトルに変換し、文脈的特徴量として利用した。

2.3 RoBERTa

近年、商品レビューの感情分析には BERT[8]や RoBERTa などの言語モデルが広く利用されている。RoBERTa は、BERT モデルの学習手法を改良した事前学習済み言語モデルであり[9], 本研究ではレビュー文の感情スコア (ポジティブ/ネガティブ/中立) 抽出に使用した。

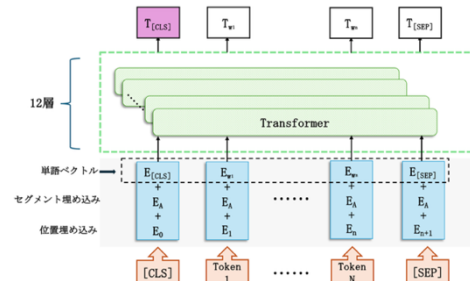


図1 Transformer モデルの構造

2.4 BERTopic

BERTopic は、事前学習済みの埋め込みモデルと階層型クラスタリングを組み合わせることで、高品質なトピック抽出を実現する手法である[10]. 本研究では、レビュー文に含まれる潜在的なトピック構造を抽出し、細粒度スコア予測のための追加的特徴量として活用した。

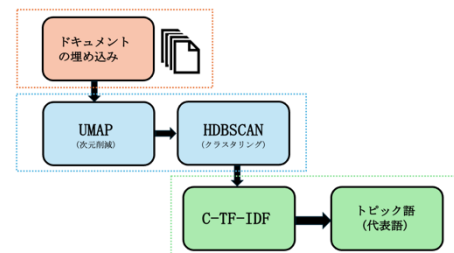


図2 BERTopic によるトピック抽出のプロセス

2.5 XGBoost

XGBoost (Extreme Gradient Boosting) は、高速かつ高精度な勾配ブースティング決定木アルゴリズムであり、さまざまな回帰・分類タスクで広く利用されている [11]。

本研究では、RoBERTa 感情スコア、SBERT ベクトル、BERTopic トピック確率を入力とし、ChatGPT によって付与した細粒度感情スコアの予測に用いた。

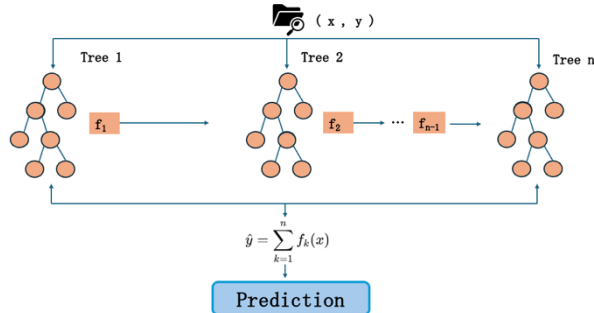


図3 XGBoost モデルの予測プロセス

3. 提案手法

3.1 提案の着想

従来のレビュー分析では、「ポジティブ/ネガティブ」の二値分類や星評価平均値に基づく傾向把握が中心であり、とりわけ星評価 3 レビュー内部の微細な感情差異を定量化する手法は確立されていない。

しかし、星評価 3 レビューには実際には幅広い感情が含まれており、そのまま扱うと推薦精度やランキング性能の低下を招く可能性がある。

本研究では、星評価 3 レビューを 3.0~3.8 の連続値スコアに細分化し、推薦やランキングに活用可能な細粒度感情スコアの生成を試みた。既存の星評価はレビュー文に含まれる文脈依存的な感情表現を十分に反映できないため、大規模言語モデル ChatGPT の言語理解能力を活用し、レビュー本文の内容に基づいて一貫性のある擬似的な細粒度スコアを付与するアプローチを採用した。

これにより、レビュー文の意味的・感情的特徴量と連続的なターゲット変数との関係性を学習し、レビュー内部の評価傾向をより高精度に捉えることを目指す。RoBERTa による感情スコア、SBERT による文ベクトル、BERTopic によるトピック確率分布を特徴量とし、ChatGPT により付与した弱ラベルを用いて XGBoost 回帰モデルを学習した。



図4 提案手法の流れ

3.2 データ準備と特徴量設計

本研究で用いた Amazon 商品レビュー・データセットは、Amazon 公式の公開データセット (Amazon Customer Reviews Dataset) を利用した。その中から星評価 3 のレビュー 3,412 件を対象データとした。前処理として、重複データや 10 文字未満の極端に短いレビューを除去し、HTML タグや特殊文字を削除した上で、テキスト正規化を行った。細粒度感情スコアリングのため、レビュー文から以下の特徴量を抽出した。

まず、レビュー文の意味的な特徴を表現するため、Sentence-BERT (SBERT) を用いて各文書 s_i を固定次元の文ベクトル $\mathbf{u}_i \in \mathbb{R}^d$ ($d=768$) に変換した。

$$\mathbf{u}_i = \text{SBERT}(s_i)$$

SBERT はコサイン類似度を最大化するタスクで学習されており、文間の意味的な類似性を高精度に表現できる。

次に、レビュー文の感情的傾向を定量化するため、事前学習済みの RoBERTa モデルを用いて感情分類を行った。Softmax 関数により肯定・中立・否定の確率 $\mathbf{p}=[p_{\text{neg}}, p_{\text{neu}}, p_{\text{pos}}]$ を出力し、これらを $[-1, 1]$ 区間の感情スコアに変換した。

$$\text{SentimentScore} = -1 \times p_{\text{neg}} + 0 \times p_{\text{neu}} + 1 \times p_{\text{pos}}$$

また、レビュー文の潜在的な話題構造を補助的に捉えるため、BERTopic によりトピッククラスタリングを行った。SBERT ベクトル \mathbf{u}_i に基づき、各文書がトピック t_k に属する確率分布 $P(t_k / s_i)$ を取得した。これにより、レビューが関連する話題情報を特徴量として活用可能とした。一方、教師あり学習のターゲット変数としては、レビュー文の内容からより細かな感情傾向を捉えることを目的とし、ChatGPT を用いて擬似的な細粒度スコア (3.0~4.0 の範囲で 0.2 刻み) を付与した。具体的には、各レビュー本文を ChatGPT に入力し、元の星評価や事前に計算した各種スコアなどの情報は一切含めず、テキスト内容のみを基にスコア付けを行った。

こうして得られたスコアを、従来の離散的な星評価に比べ、より連続的かつ柔軟な感情傾向の学習を可能にする弱ラベルとして活用した。

3.3 XGBoost によるスコア予測

整備した特徴量 (RoBERTa 感情スコア、SBERT ベクトル、BERTopic トピック確率分布) と ChatGPT 生成の細粒度スコアを用いて、XGBoost 回帰モデルを学習した。

ここで y_i はターゲットである ChatGPT による細粒度スコア、 $\hat{y}_i^{(t-1)}$ は $t-1$ 回目までの予測値、 f_t は t 回目の決定木、 $l(\cdot)$ には平均二乗誤差 (MSE) を採用した。モデルの性能は決定係数 R^2 および平均絶対誤差 (MAE) を用いて評価した。

3.4 特徴量分析

学習済みの XGBoost モデルに対して特徴量重要度分析を行い、どの特徴量が細粒度スコア予測に寄与しているかを可視化した。特に 3.6~3.8 帯に強く関連するトピック特徴量を重点的に分析し、レビュー内部の肯定的傾向との関係性を明らかにした。

また補助的な分析として、トピック確率分布に対してリッジ回帰を適用し、線形モデルの係数から各トピックの細粒度スコアに対する正負の影響方向を可視化した。これにより肯定的スコアと強く関連するトピック (例: 音質, デザイン), および否定的スコアと関連するトピック (例: 返品, 初期不良) を把握することができた。

本結果は、レビュー内部に潜在する評価要因と満足度傾向の関係をより明確に示すものである。

4. 実験結果

4.1 RoBERTa による感情スコア分布

本研究で提案した細粒度感情スコアリングおよび回帰予測モデルの効果を検証するため、Amazon レビューに対する分析結果を以下に示す。

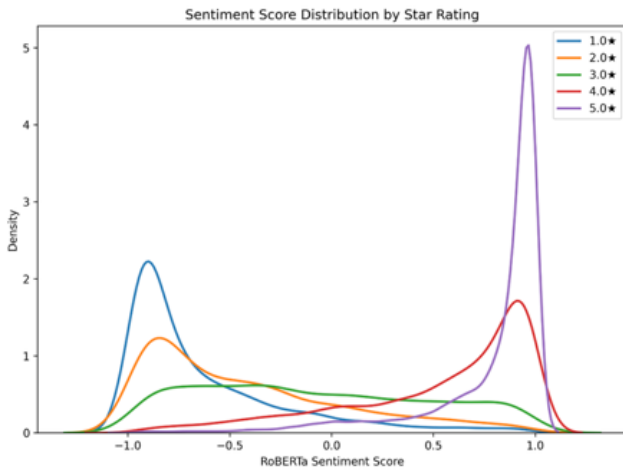


図 5 RoBERTa による感情スコア

図 5 は各星評価における RoBERTa 感情スコアの密度分布である。星評価 3 評価 (緑線) は、他の星評価と比較して広範な感情分布を示しており、特にスコアが正の領域では星評価 4 (赤線) と大きく重なっている。このことから、星評価 3 レビューの中には肯定的な内容を含む「隠れ高評価」が多く存在することが示唆される。

4.2 細粒度スコアの予測結果

図 6 および図 7 は、ChatGPT により付与された True ラベルと、XGBoost モデルによる Predicted ラベルの分布比較である。True ラベル (図 6) では 3.6 および 3.8 にピークが見られ、一部のレビューがポジティブ寄りとして高スコア付与されている。Predicted ラベル (図 7) でも同様の傾向

が確認されており、本モデルがポジティブ寄りのレビュー予測に成功していることがわかる。これにより、提案手法が星評価 3 レビュー内部の細かな感情傾向を定量的に学習・予測できていることが確認できる。

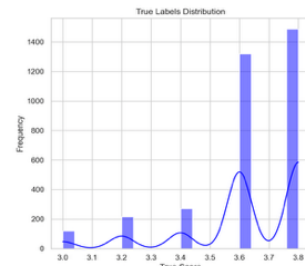


図 6 True ラベル

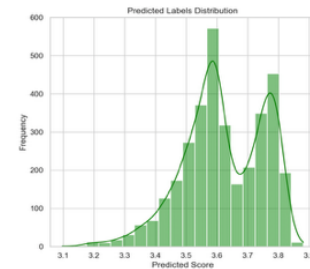


図 7 Predicted ラベル

XGBoost モデルの予測性能としては、決定係数 $R^2=0.665$ 、平均絶対誤差 (MAE) =0.093、二乗平均平方根誤差 (RMSE) =0.108 を達成した。これにより、細粒度スコアの回帰予測が高い精度で実現できていることが示された。

4.3 XGBoost における特徴量の重要度

図 8 に XGBoost モデルにおける特徴量の重要度を示す。

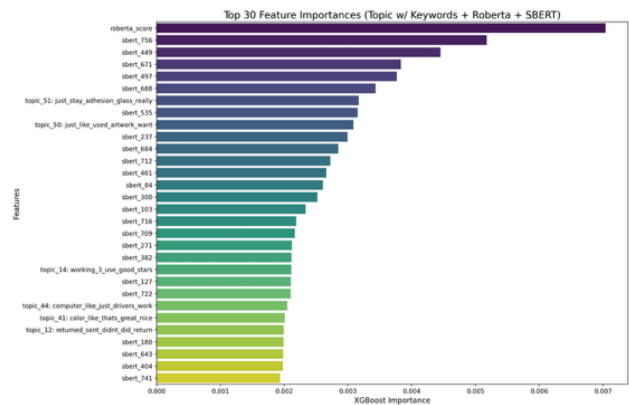


図 8 XGBoost モデルにおける特徴量の重要度

RoBERTa スコアが最も高い重要度を示したほか、BERTopic により抽出された複数のトピック特徴量が上位に入り、レビューの話題的側面が細粒度スコア予測に大きく寄与していることが示された。また、SBERT ベクトルも多数ランクインしており、レビュー文全体の文脈的な意味情報が有効な補助情報として機能していることが確認できる。

4.4 トピックと細粒度スコアの関係性分析

Ridge 回帰に基づくトピック特徴量の線形係数 (β) を図9に示す。係数が正のトピックはスコアが高い (ポジティブ寄り) 傾向と、係数が負のトピックはスコアが低い (ネガティブ寄り) 傾向と強い関連性を示している。特に「sound_quality」や「操作性」といったポジティブ評価と結びつくトピックが正の係数を有し、一方で「返品」や「故障」など消費者の不満を反映するトピックが負の係数を示していることが確認された。

本分析により、レビュー中に潜在するトピックごとの定量的な影響方向と影響度を明示的に可視化することが可能となった。これにより、従来は定性的に捉えられていた消費者の評価要因が、数理モデルを通じて体系的に理解できる枠組みが提供される。特に EC サイト運営や製品設計の現場においては、ポジティブな影響を有するトピックの強化、およびネガティブなトピックの低減に資する具体的な改善指針を導出する上で有用な知見が得られると考えられる。

さらに、本手法の有用性は、レビュー活用的高度化という観点でも示唆的である。例えば、マーケティング施策や商品改善の優先度設定において、各トピックの影響度を踏まえたデータ駆動型の意思決定が可能となる。今後は、より多様な製品カテゴリや多言語レビューコーパスへの適用を通じて、モデルの汎用性と実用性を検証し、EC サイトにおけるレビュー活用戦略の高度化に貢献することが期待される。

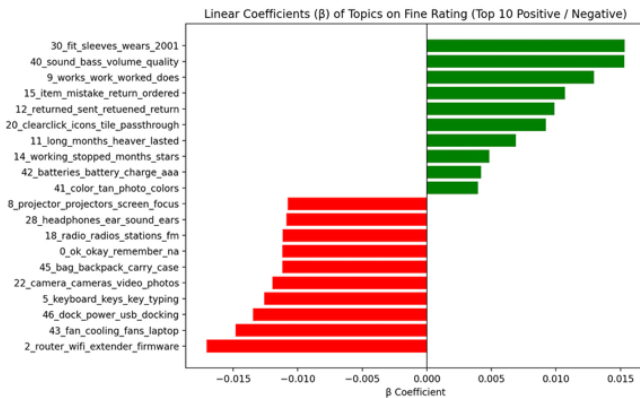


図9 トピック特徴量と細粒度スコアとの関係性

5. 考察

本研究の結果より、星評価3のレビューには実際には星評価4相当のポジティブな内容を含むものが多数存在することが確認された。RoBERTa の感情スコアや細粒度スコアの分布において、3.6~3.8帯に集中するレビューが見られ、現行の整数ベースの星評価ではユーザーの実際の満足度を十分に反映できていないことが示唆された。

この点から、EC サイトの設計においては「3.5星」などの中間的な評価選択肢を導入し、星評価の粒度を部分的に高めることが有効である可能性がある。これにより、レビュー

投稿時の表現精度と、推薦・ランキングの精度向上、ユーザーの購買判断支援が期待される。

6. 結論と今後の課題

本研究では、EC サイトにおける星評価3のレビューを対象に、RoBERTaによる感情スコア、SBERTベクトル、BERTopicトピック確率を組み合わせた細粒度スコアリング手法を提案した。XGBoost回帰モデルを用いて、星評価の内部構造をより精緻に捉える可能性を示した。

今後は、他の星評価帯への適用や人手付与ラベルとの比較検証、実サービスへの展開を通じて、本手法の汎用性と実用性をさらに高めていきたい。

参考文献

- [1] Chevalier, J. A., & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research*, 43(3), (pp. 345–354).
- [2] Liu, Y. (2014). The impact of word of mouth on sales: Online movie reviews. *Journal of Marketing*, 70(3), (pp. 74–89).
- [3] Jannach, D., & Ludewig, M. (2017). When recurrent neural networks meet the neighborhood for session-based recommendation. *Proceedings of the Eleventh ACM Conference on Recommender Systems (RecSys 2017)*, (pp. 306–310).
- [4] Zhang, Y., et al. (2021). Learning to Rank with Multiple Relevance Criteria via Data Augmentation. *Proceedings of the Web Conference 2021 (WWW '21)*, (pp. 2290–2301).
- [5] Guo, H., et al. (2020). Multi-Aspect Sentiment Analysis with RoBERTa and Data Augmentation. *Proceedings of COLING 2020*, (pp. 6479–6489).
- [6] Taywade, S., et al. (2023). Sentiment Classification of Twitter Data using RoBERTa Model: A Comprehensive Survey and Evaluation. *Procedia Computer Science*, 218, (pp. 1995–2002).
- [7] Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. *Proceedings of EMNLP-IJCNLP 2019*, (pp. 3982–3992).
- [8] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, (pp. 4171–4186).
- [9] Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). RoBERTa: A Robustly Optimized BERT Pretraining Approach. *arXiv preprint arXiv:1907.11692*.
- [10] Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF strategy. *arXiv preprint arXiv:2203.05794*.
- [11] Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, (pp. 785–794).