

# 食品 EC サイト投稿コメントのベクトル化による 消費者の商品評価推定モデルの構築

○佐藤 瑠一<sup>1</sup>, 赤木 徳顕<sup>2</sup>, 桑原 豊史<sup>3</sup>

**キーワード:** 大規模言語モデル, 埋め込みベクトル, 消費者コメント, 多変量ロジスティック回帰

## 1. はじめに

商品やサービスに対する消費者の知覚を測定する手法としては、アンケート調査等で順序尺度を用いることが一般的である。しかし、回答者には多くの質問を理解し回答する負担が生じ、調査側も作問・分析時に回答者毎に異なる尺度感覚を調整する必要があるといった課題が存在する。

このような消費者の知覚や行動の分析手法は、生成 AI 技術の急速な発展によって多様化している。自然言語のベクトル化技術はその 1 つであり、単語や文章の意味を高次元ベクトルとして定量的に表現することが可能となった[1]。

本研究では、アンケート調査に依存せずに消費者の知覚を定量評価するためのアプローチを探索した。具体的には、食品 EC サイト利用者が投稿したコメントデータのベクトル傾向を機械学習し、各コメントがどのレビュースコア尺度に相当するかを分類するモデルを構築・評価した。

## 2. 方法

本研究では、株式会社ビビッドガーデンが運営する食品 EC サイト「食べチョク」に投稿された商品評価コメント (n=15,386) と、それに対応するレビュースコア (5 段階評価、5 が最良) を分類モデル構築の学習に使用した[2]。

まず、食べチョクに投稿されたユーザーコメントを、OpenAI が提供する text-embedding-3-small を用いて 1,536 次元のベクトルに変換した[3]。このベクトルを説明変数、レビュースコアを目的変数にした多変量ロジスティック回帰モデルを構築し、コメントデータから直接商品評価を推定する分類モデルを実装した。モデル評価には AUC を使用し、k 分割交差検証 (k=5) で汎化性能も併せて検証した。

次に、モデルの精度を高めるため、分類に不要なベクトルを削減した。具体的には、ベクトルごとに個別のロジスティック回帰を実施し、単体での分類性能を AUC として算出した。その後、0.02 刻みで閾値を設定し、分類性能の低いベクトルを順次削減した。その後、削減後のベクトル群を用いて多変量ロジスティック回帰モデルを構築し、各レビュースコアの分類性能を AUC で評価した。

最後に、構築したモデルの分類能力を検証するため、ChatGPT を用いて生成したダミーコメントをモデルに適用

した。モデルが分類したコメントにおける頻出単語を抽出し、ワードクラウドを用いて視覚化を行った。

## 3. 結果

分析に使用したコメントのレビュースコア別件数は、最良評価である 5 が最も多く、8,065 件を占めた。一方、最も少なかったのは Score\_2 で 129 件であった (図 1)。

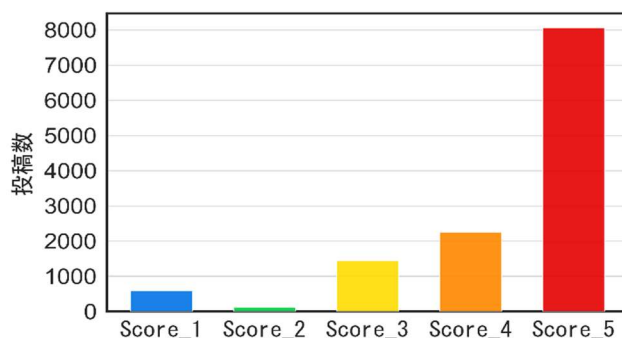


図 1 レビュースコア別の投稿件数

まず、多変量ロジスティック回帰によって得られたモデルの分類精度を以下に示す (図 2)。いずれのレビュースコアの推定精度も AUC は 0.6 を上回っており、コメントを高い精度で分類できていることが確認された。特に、Score\_1 にコメントを分類する精度が最も高かった。また、k 分割交差検証を用いて汎化性能を検証した結果、学習データが少ない Score\_2 のみ標準偏差に大きなばらつきが見られた。しかし、全ての検証試行で AUC 0.6 以上を維持しており、安定した分類性能があることが示唆された。

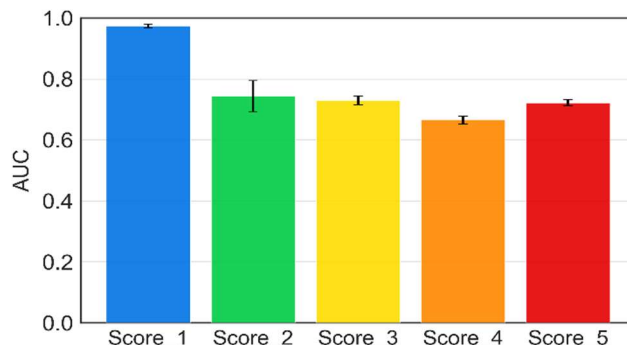


図 2 多変量ロジスティック回帰モデルの分類精度

1 岩手大学大学院総合科学研究科  
2 開志専門職大学

3 株式会社ビビッドガーデン

次に、単体での分類性能が低いベクトルを削減した後のベクトル群を用いた分類モデルの性能評価を以下に示す(図3)。結果、各レビュースコアの最適な閾値は、Score\_1で0.3、Score\_2で0.62、Score\_3で0.52、Score\_4で0.54、Score\_5で0.5だった。Score\_2のみベクトル削減による分類性能の改善が見られたものの、それ以外のスコアでは改善が見られなかった。また、Score\_1を除き、ベクトル削減がある一定の閾値を行き過ぎると、レビュースコアの分類能力が急激に低下し、分類能力を失うことが示唆された。

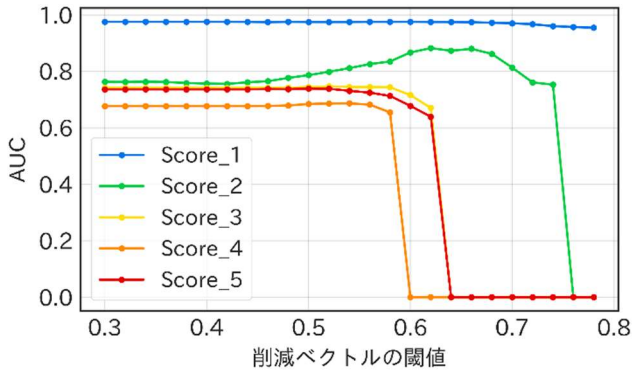


図3 ベクトルを閾値で削減して AUC 改善

最後に、構築したモデルの分類能力を検証するため、ChatGPT を用いて生成したダミーコメントにモデルを適用した。そのプロンプトには「食品 EC サイトで購入した桃が到着したことに對するコメントを5段階評価に對して生成」を指定した。結果、ダミーコメントを適切にスコアリングしていることが確認された(図4、図5)。これにより、ユーザーから実際に入力されたレビュースコアと同等の意味合いの尺度に分類できていることが示された。



図4 分類したダミーコメントの頻出単語 (Score\_1)



図5 分類したダミーコメントの頻出単語 (Score\_5)

#### 4. おわりに

本研究では、食べチョクサイトに投稿されたユーザーコメントをベクトル化し、コメントデータから直接レビュースコアを推定する分類モデルを実装することで、人手による判定に依存せず、消費者知覚を順序尺度として定量評価するアプローチを模索した。

その結果、従来の人手によるスコア入力と同等の分類を、コメントデータのみから高精度で実施できるモデルの構築に成功した。これにより、事前に学習モデルを用意することで、その都度消費者のアンケート入力に頼ることなく、テキストデータのみから順序尺度を推定する手法の可能性を示すことができた。本手法の応用例として、インタビュー音声の文字起こしデータに適用し、既存の5段階評価手法と連携させることで、新たな分析の可能性が広がると考えられる。

一方で、本研究と同様に AI を活用した手法との精度比較という懸念が残されている。例えば、ChatGPT などの LLM に「顧客コメントを直接5段階評価して」と命令を与えることで、本研究と似たような回答を得ることができる。ただ、この手法にも、LLM が複雑な否定表現の解釈を苦手とする(日本の首都じゃない都市と聞いても、「日本×首都」として確率の高い東京を回答してしまう)、分析過程がブラックボックスであるという課題がある。そのため本研究の方が否定文対応・解釈の容易さの観点から優位性があると考える。しかし、効率性や適用性の観点から、どちらを採用すべきかについてはさらなる検討が必要である。

#### 謝辞

本研究を遂行するにあたり、株式会社ビビッドガーデン様より多大なるご協力を賜り、分析データをご提供いただきました。研究の基盤となる貴重なデータをご提供いただきましたことに深く感謝申し上げます。本研究が成り立ったのは、同社のご支援とご理解あつてのことと心より感謝いたします。

#### 参考文献

- [1]大山百々勢, 横井祥, 下平英寿 (2022): 単語ベクトルの長さは意味の強さを表す. 言語処理学会第 28 回年次大会発表論文集, 1050-1055.
- [2]株式会社ビビッドガーデン: 食べチョク ホームページ <<https://www.tabechoku.com/>>2025.1.31 参照
- [3]OpenAI: Vector embeddings - OpenAI API <<https://platform.openai.com/docs/guides/embeddings>> 2025.1.31 参照