

Amazon における消費者レビューの削除について¹⁾

黒 田 敏 史

要約

本稿では、Amazon がどのようなレビューを削除しているのかを、Cox 比例ハザードモデルを用いて分析した結果を紹介する。分析結果から、削除されやすいレビューには一定の特徴が存在するものの、消費者が文章を目視して動機づけられたレビューを識別することは困難であり、自助的な努力には限界があることが示唆された。

1. はじめに

消費者が投稿するレビューは、実際に商品を使用した経験を他者に伝えることで、生産者が直接届けることのできない情報を補完する役割を果たしている。一方で、レビューの中には商品の販売者が高評価を得ることを意図して評価者に真の評価ではない高い評価をするよう動機づけた虚偽の情報が存在することも知られている。

Amazon は利用規約において、動機づけられたレビューを禁止し、発見した場合には削除すると定めている。したがって、投稿後に削除されたレビューには、Amazon が偽レビューと認定したものが含まれる。本研究では、Amazon の商品ページを定期的に観察し、投稿レビューのテキストやスコア、ならびに掲載状況を記録したデータを用いて、Amazon がどのようなレビューを削除しているのかを分析する。

2. モデルと識別戦略

本研究では、レビューが投稿されてから削除されるまでの期間を「レビューの生存期間」と定義し、Cox 比例ハザードモデルを用いて削除までの期間を推定した。データ取得期間中に削除されなかったレビューは打ち切りデータとして扱った。分析期間のうち、2022年に投稿され 2024 年に削除が確認されたレビューは、Amazon が十分な判断期間を経てフェイクレビューと判断したものとみなせる。

Amazonにおける消費者レビューの削除について

各レビューに関して観察される共変量はレビュー スコアが 5 点であるダミー, レビューに含まれる画像の数, レビュー投稿者が当該商品を Amazon で購入したかのダミー, 販売前に投稿されたレビューのダミー, レビューが役に立つと評価された件数, 商品の販売者の数, 商品の販売者の所在地割合 (中国, 日本または中国以外, 不明), 当該商品を販売した者の数, 商品が発売されてからレビューが投稿するまでの期間, Amazon による販売があるか否かのダミー, Amazon のみが販売者であるダミー, 商品固定効果, レビュー文字列を表す多次元のベクトルである。モデルに含まれる変数が多いため, Simon et al. (2011) によるペナルティ付き変数選択モデルを利用して変数選択を行った後, 選択された変数を用いてモデルの一致推定量を求めた。

3. 利用したデータ

本研究では大磯他 (2023) と同様に日時で Amazon.co.jp のヘッドホンカテゴリのレビュー ページを取得したデータを用いた。取得したレビュー数は 613,410 個である。レビュー コメントは Reimers (2019) による Sentence-BERT を用いて 768 次元のベクトルへと数量化した。

4. 推定結果

以下の表 1 が COX 比例ハザードモデルを用いてレビューの生存期間を推定した結果である。

表 1 COX 比例ハザードモデルによるレビュー生存期間分析

変数	推定値	標準誤差
スコア 5	0.277	0.014***
画像の数	0.106	0.006***
発売前レビュー	0.01	0.241
販売者: 中国所在	2.103	0.037***
販売者: その他の外国	0.717	0.115***
販売者: 所在地不明	0.273	0.036***
発売からレビュー投稿までの期間	0.035	0.000***
売り手が Amazon	0.106	0.058
売り手が Amazon outlet	0.048	0.055
標本数	613,410	
AIC	1,028,276	
BIC	1,043,125	

* p<0.05 ** p<0.01 *** p<0.001

表2 削除されたレビュー文章の例

削除されたレビュー文章

削除されたレビュー文章
この価格でとてもいいと思います
とてもいい商品で、気に入っています。
見た目も良く、とても気に入っています。
この値段で、この品質は素晴らしいです！
この価格で買えるのは素晴らしいことだと思います
使いやすくデザインもよく、値段も手頃で、とても気に入っています。
コストパフォーマンスが高くて大満足です。
使い心地がいい、とてもいい
とても良いです、満足しています
お値段の割によい品質で大変満足しております。

表3 削除されていないレビュー文章の例

削除されていないレビュー文章

削除されていないレビュー文章
この価格でこの品質には、とても満足しています。
この値段でこの品質、とても満足しています。
この価格でこの品質でとても満足です。
この値段で、とても良いと思います
値段の割には、とても良いです
この価格にしては、とても良いと思います。
この値段で、この品質は非常に満足です。
この価格でこの質、素晴らしいと思います。
この値段でとても良いです
コストパフォーマンスはとても良いです。

モデルの推定結果から、最高評価である5点のレビュー、画像の多いレビュー、中国やその他の外国の売り手の多い製品のレビュー、売り手の国籍が不明なレビュー、発売されてから書かれるまでの時間の長いレビューにおいて短期間で削除される結果となっている。また、表には不掲載であるが、レビューベクトルは768次元中342次元が有意な係数を持っており、本文の記載内容も削除のされやすさと関連していることがわかった。

続いて、どのような文章のレビューがより削除されやすいかを目視するため、有意な影響を持つ文章ベクトルを2次元の主成分に次元削減し、二次元空間の中心からの距離が高く、評価が5点レビューのうち、削除されたレビューと削除されなかったレビューからそれぞれ文章を10個抽出したのが以下の表2、3である。

いずれの表のレビューも商品の価格や品質を高く評価する短い文章になっており、平均的なレビューとは異質ではある。削除されたレビューよりも削除されていないレビューの方が文章の均質性が高いように見えるが、どのレビューが削除されたレビューであるかを閲覧者

Amazon における消費者レビューの削除について

が判別することは極めて困難であろう。

このことは、Amazon のレビュー削除基準には消費者が文字列として把握できる情報以外の様々な要素から動機づけられたレビューを検出しており、消費者にとってどのレビューが動機づけられたレビューであるのかを目視で判断する事が難しく、偽レビューに対応するための消費者の努力には限界があることを示唆する。

注

1) 本研究は、2024 年度の東京経済大学個人研究助成費（研究番号 24-09）を受けた研究成果である。

参考文献

- 大磯 一, 坂口 洋英, 黒田 敏史, 依田 高典, Effect of the Product Order and the Product Review on the Product Choice, 行動経済学, 2023, 16 卷, Special_issue 号, p. S13–S16.
- Simon, N., J. H. Friedman, T. Hastie, and R. Tibshirani, Regularization Paths for Cox's Proportional Hazards Model via Coordinate Descent. Journal of Statistical Software, 2011, 39 (5), 1–13.
- Reimers, N. and I. Gurevych, Sentence-BERT: Sentence embeddings using siamese BERT-networks, 2019, arXiv.