

商品レビュー文の有用性推定に関する考察

A study of estimation of the review helpfulness

柴田 知親^{1*} 伊東 栄典^{2†}
Tomochika Shibata¹ and Eisuke Ito²

¹ 九州大学大学院システム情報科学府

¹ Graduate School of ISEE, Kyushu University

² 九州大学情報基盤研究開発センター

² Research Institute for IT, Kyushu University

Abstract: Product reviews can be helpful for consumers to finalize their purchasing decisions. The advertisement contains sentences intended by the seller. Meanwhile, the product review seems to be an honest impression of the consumer, so it is easier to trust. However, there are some problems such as fake reviews and guerilla marketing in recent years. Therefore, in many e-commerce sites, each reviews have “Helpful” and “Not Helpful” buttons for consumers to evaluate the helpfulness of the review. In this paper, we consider the estimation model of review helpfulness for the review sentences of Rakuten Ichiba (rakuten.co.jp).

1 はじめに

商品レビューは、消費者が購入時の判断に参考にするため、オンラインショッピングに大きな影響を与える。広告には売手の意図した文章が含まれるのに対し、商品レビューは購入した客の正直な感想と思われるため信頼しやすい。そのため多くのECサイトでは、情報提供の一つとして利用者による商品レビューを掲載している。

しかし近年、ヤラセやステマなどのレビューの悪用 [J-Cast 15] や利用者の増加に伴う役に立たないノイズレビューの増加により、真に有用なレビューが埋もれる問題がある。この問題に対し、図3のように、利用者「このレビューは参考になりましたか?」とレビューを評価させる有用性評価が行われている。レビューごとに、「参考になった」、「参考にならなかった」などのボタンが設置されており、利用者は任意のレビューに対してその有用性を評価できる。また、「参考になった」の回数が多いレビューに対してはその人数がレビューとともに掲載される。この人数を利用してレビューを並びかえることで、有用性の高いレビューを検索できるシステムなども多く存在する。この有用性評価は、Amazon.comに毎年27億ドルのさらなる売上をもたらしているという報告もある。 [Spool 09]

既に多くのレビューが存在する場合、新たに投稿された有用なレビューが過去のものと同じように評価される可能性は低い。投稿日時に関係なく、レビューそのものの有用性を推定できれば、より多くの有用な情報を消費者および出品者に提供できる。本論文では、我々がこれまでに研究してきた [林 15], [Kusakari 18] 楽天市場のレビュー [楽天] を対象に、レビューの文章から、そのレビューの有用性推定について考察する。

本文の構成を述べる。2節では、商品レビューの有用性推定についての関連研究の概要を述べる。3節では、本研究で使用したレビューデータセットの概要およびその特徴、有用性スコアの算出方法について説明する。4節では、用いる特徴量の説明および評価方法と、予測結果を述べる。5節では結果についての考察を述べる。最後に6節でまとめと今後の課題を述べる。

2 関連研究

[Martin 14] では、Amazon, Trip Adviser, Yelp の3つのサイトにおけるレビューに対して、有用性の推定を行っている。特徴量として、単語のTF-IDFに加えて、テキストの長さ、レビューのレーティング、感情辞書を用いて作成した感情ベクトルなどを与えている。分類および回帰を行い、感情ベースの特徴を用いた方法が優れていることを示している。

[Yang 15] では、Amazon Review Dataset [Mcauley 13] のうち、4つのカテゴリ (Books, Home(home and kitchen), Outdoors, Electronics) からなる696,696件のサブセッ

*連絡先：九州大学大学院システム情報科学府
〒 819-0395 福岡県福岡市西区元岡 744
E-mail: t.shibata.130@s.kyushu-u.ac.jp

†連絡先：九州大学情報基盤研究開発センター
〒 819-0395 福岡県福岡市西区元岡 744
E-mail: ito.eisuke.523@m.kyushu-u.ac.jp

トを使用している。モデルの転用しやすさを考慮し、レビュー文章のみから得られる情報を特徴量としている。[Martin 14]と同様に複数の感情辞書を用いて、SVRモデルによる回帰分析を行っている。予測する有用性の値は X/Y としている。 X は各レビューの「参考になった」の数、 Y は各レビューの「参考になった」の数と「参考にならなかった」の数の合計である。また、 Y が非常に小さく、分析に用いた Y が5以上のデータは全体の20%未満しかないなど、偏っている点を指摘している。さらに、人手で0から100の点をつけた400件のレビューデータセットを作成し、同様の分析を行っている。このデータセットに対してはより相関性の高い予測ができており、 X/Y をレビューの評価値とするのは、人間の主観性により有用性を上手く近似出来ないのではと推測している。

3 データセット

本研究が対象とするデータは、楽天株式会社が国立情報学研究所の情報学研究リポジトリ [NII] を通じて研究者に提供している楽天市場データセット [楽天] である。このデータセットには、楽天市場で販売される商品の情報と、商品へのレビューが含まれている。利用者から投稿された商品へのレビューは約6400万件ある。期間は、2010年1月から2012年12月までの3年間(36ヶ月)である。

楽天市場データセットには、レビューの有用性を示す「参考になった」の数が含まれていない。そのため、データセット内の商品URLを利用して、楽天のサイトから再度レビューデータを取得した。レビュー取得のためのプログラムはPython言語で作成した。

3.1 楽天レビューデータ

本研究で使用したデータについて表1に示す。収集期間は2018年10月11日から2018年10月23日である。提供されているデータのうち、収集期間までに商品ページが削除されたものは含まれていない。また、各商品のレビューは「参考になった」の数(以後 $vote$)が多い順に最大1500件取得した。

表 1: 楽天レビューデータ数

項目	値
商品数	159,031
レビュー数	12,414,533

3.2 各ジャンルの特徴

本研究では、各ジャンルごとにレビュー件数が1000以上ある商品のレビューを対象に、レビュー数上位10個のジャンルについて分析した。表2に分析した各ジャンルのレビューの特徴を示す。 $vote > 0$ の割合を見ると、多くのジャンルで9割以上のレビューが全く評価されていないことが分かる。

3.3 有用性スコア

目的変数となる有用性スコアについて説明する。表2にある通り、未評価のレビューが全体の9割以上を占める。未評価の中には、真に無価値のものと殆ど読まれていない有用なレビューとが混在している可能性がある。そのため、本研究では、 $vote > 0$ のレビューのみを用いて分析した。

各商品を I_i とし、各レビューの $vote$ を $v_{i,j}$ とする。「参考にならなかった」の数は取得できていないため、[Yang 15]と同様に有用性スコアをつけることが出来ない。そこで、以下の式1で表される値 $y_{i,j}$ を本研究における有用性スコアとした。

$$y_{i,j} = \log \left(\frac{v_{i,j}}{\sum_j v_{i,j}} \right) \quad (1)$$

また、図1に示す、ジャンル「インテリア・寝具・収納」における正規化された有用性スコアのヒストグラムを見ると、正規分布になっていることが確認できる。他のジャンルについても同様の結果を得た。

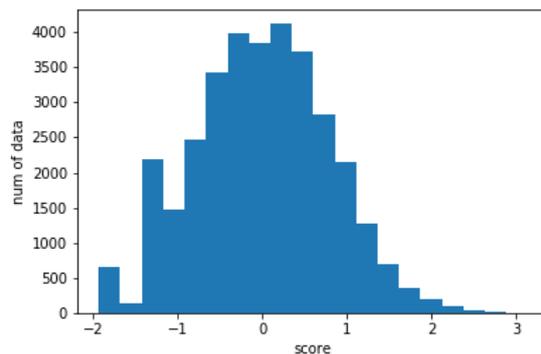


図 1: 有用性スコアのヒストグラム

表 2: 各ジャンルの特徴

ジャンル	商品数	レビュー数	vote > 0	vote > 0 の割合
インテリア・寝具・収納	315	435,796	33,709	7.7%
日用品雑貨・文房具・手芸	239	331,847	29,161	8.8%
食品	224	314,702	43,218	13.7%
バッグ・小物・ブランド雑貨	222	306,883	22,386	7.3%
キッズ・ベビー・マタニティ	205	284,793	20,878	7.3%
レディースファッション	152	206,861	25,190	12.2%
靴	137	189,019	16,510	8.7%
家電	103	145,947	14,101	9.7%
メンズファッション	88	118,999	7,116	6.0%
スポーツ・アウトドア	83	113,742	7,105	6.2%

4 実験

4.1 手法

[Martin 14], [Yang 15] と同様に SVR モデルによる回帰を行う。目的変数は 3.3 節の式 1 を正規化した値とする。また、特徴量として、単語の TF-IDF を用いた UGR 特徴 (Uni-Gram 特徴) と、文字数や改行数、内容詞の数など構造的特徴を加えた STR 特徴 (Structural 特徴) を比較する。表 2 の各ジャンルについて、実測値と予測値の相関係数で評価する。評価値は 5 分割交差検定による平均とする。

4.2 結果

4.1 節で述べた手法を表 2 に示す 10 個のジャンルのレビュー群に適用した。その結果を表 3 に示す。10 個のジャンル全てにおいて、STR 特徴を用いた方が相関係数が高い結果となった。また、相関係数はおおむね 0.5 となり、[Yang 15] の X/Y を有用性スコアとした際の結果と同程度の相関係数を得た。

表 3: 各ジャンルにおける予測結果

ジャンル	UGR	STR
インテリア・寝具・収納	0.45937	0.48739
日用品雑貨・文房具・手芸	0.54067	0.57900
食品	0.46011	0.48367
バッグ・小物・ブランド雑貨	0.45170	0.45766
キッズ・ベビー・マタニティ	0.63492	0.64014
レディースファッション	0.53990	0.54029
靴	0.39732	0.40389
家電	0.70409	0.71388
メンズファッション	0.49263	0.49711
スポーツ・アウトドア	0.71729	0.72148

5 考察

表 3 を見ると、「家電」、「スポーツ・アウトドア」においては相関係数が 0.7 以上となり、比較的高い結果を得た。「家電」については、商品の機能や特徴をレビューに含むことで、客観的な評価を得やすく、比較的推定し易いと考えられる。「スポーツ・アウトドア」については、7,105 件のレビューのうち 1,500 件が一つの商品のものであったため、この商品のレビューが大きく影響したと考えられる。

本研究では、ジャンルごとにデータを分割し、それぞれで SVR モデルを学習させ評価した。しかし、楽天市場のジャンルは階層的に分けられており、図 2 のように多くのサブジャンルが存在する。商品のジャンルが異なれば当然評価されるレビューの内容も異なる。そのため、今回分析したジャンルをなるべく類似種の商品からなる部分集合に分割することで、より高い精度で推定できる可能性がある。コンテンツの階層的分割による閲覧回数の回帰分析については、我々の研究 [柴田 18] で一定の精度向上を確認している。

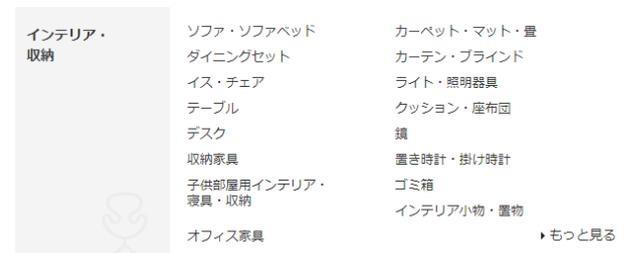


図 2: インテリアのサブジャンル

6 おわりに

本研究では、楽天市場の商品レビューを対象に、その有用性の推定について考察した。関連研究である

[Martin 14], [Yang 15] を参考に, SVR モデルで有用性推定を行った. データの特性上, 有用性の定義は異なるが, 関連研究と同程度の推定精度を得た.

今回はモデルの転用しやすさを考慮してレビュー本文から得られる特徴のみを用いて推定したところ, 全体的な精度は決して高くない. ジャンルをさらに細かく分割することで, 推定精度の向上を図れる. しかし, ジャンルで分割している時点で, レビュー本文以外に, ジャンルという商品情報を加味している. このことから, 商品のジャンルや説明文なども含め, 商品固有の情報とレビュー本文からそのレビューの有用性推定を行う新たなモデルを検討したい. レビューを対象とすれば, 他のサイトなどにも十分転用可能である.

謝辞

国立情報学研究所情報学研究リポジトリおよび楽天株式会社に感謝します.

参考文献

[J-Cast 15] J-Cast ニュース: 「やらせ」の口コミや架空の注文なんと 11 万件超! 「楽天市場」で大掛かりなステマが発覚, <https://www.j-cast.com/2015/03/21231013.html?p=all> (2015)

[Kusakari 18] Kusakari, Nakatoh, Suzuki and Hirokawa: Detection of Over-Praised Reviews and Feature Word Extraction, *Forum on Information Technology 2018* (2018)

[Martin 14] Martin Lionel and Pu Pearl: Prediction of helpful reviews using emotions extraction, *AAAI'14 Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*, pp. 1551–1557 (2014)

[Mcauley 13] Mcauley and Julian: Hidden Factors and Hidden Topics : Understanding Rating Dimensions with Review Text, *RecSys '13 Proceedings of the 7th ACM conference on Recommender systems*, pp. 165–172 (2013)

[Spool 09] Jared M. Spool: The Magic Behind Amazon's 2.7 Billion Dollar Question, <https://articles.uie.com/magicbehindamazon/> (2009)

[Yang 15] Yang Yinfei and Bao Forrest Sheng: Semantic Analysis and Helpfulness Prediction of Text for Online Product Reviews, *Proceedings of the*

53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), pp. 38–44 (2015)

[楽天] 楽天株式会社: 楽天市場レビューデータ, https://rit.rakuten.co.jp/data_release_ja/

[NII] 国立情報学研究所: 情報学研究データリポジトリ, <https://openforum.nii.ac.jp/dsc/idr/index.html>

[柴田 18] 柴田 知親, 伊東 栄典: 回帰分析による CGM 動画再生回数推定, 第 10 回データ工学と情報マネジメントに関するフォーラム (2018)

[林 15] 林 驍, 伊東 栄典, 廣川 佐千男: 大規模 EC サイトの商品レビュー傾向分析, 第 181 回知能システム研究発表会 (2015)

★★★★★ 4

2011-01-09

商品の使いみち:実用品・普段使い 商品を使う人:自分用 購入した回数:はじめて

ずーっとほしかったので

うちはいつもインスタントを飲んでいるので、とてもよかったです。普段ちゃんとコーヒーを入れている方は物足りないかも。便利なのは確かです。カフェラテが簡単にできる！お友達に出すときもインスタントでもこれだったらOKかな？ただ、少し大きいです。あと、楽しみにしていたエスプレッソはちょっと酸味が強く、お店で飲んだ方がいいかな、、、。エコパックが高い、との問題ですが、エコパックの空き容器を捨てないで、底を切って漏斗の用にし、割安の袋入りを入れられます。これおすすめ～。

▶ [このレビューのURL](#)

29人が参考になったと回答

このレビューは参考になりましたか？

[不適切なレビューを報告する](#)

図 3: 楽天市場の商品レビュー