

レビューの信頼性を説明付けするシステム

A System for Describing Reviewer Credibility

田中 祐也 中村 伸子 土方 嘉徳 西田 正吾*

Summary. 近年、レビューサイトに書き込みを行うユーザ（レビューア）が増加し、レビューの質が多様化している。レビューサイトにはユーザからの投票結果に基づいたレビューのランキングなどの指標が存在するが、現在用意されている指標ではユーザは具体的にレビューのどのような点が優れているのかを知ることができない。本稿では、レビューの信頼性を評価する際に有効な属性と手法を明らかにし、レビューを信頼性でランク付けするとともに有効な属性を用いてレビューの信頼性を説明付けするシステムを開発する。

1 はじめに

近年、レビューサイトに書き込みを行うユーザ（レビューア）が増加し、有益な情報を発信する信頼できるレビューだけでなく、参考にならない情報しか発信しない信頼できないレビューも書き込みを行うようになってきた。我々は、信頼できるレビューを特定することで、購入の意思決定の判断材料となる有用なレビューを特定できると考える。

代表的なレビューサイトである Amazon.com では、レビューの信頼性の指標としてユーザからの投票結果に基づいたレビューのランキングが示されている。Epinions.com では、レビューが何人のユーザから信頼されているかを確認することができる。ユーザは、これらの指標によって大まかにレビューの信頼性を評価できるものの、具体的にレビューのどのような点が優れているのかを知ることができない。本研究ではレビューを信頼性順にランク付けするだけでなく、レビューの信頼性と関連のある属性の評価値も提示することによりレビューの信頼性を説明付けするシステムの開発を行う。

2 レビューの信頼性評価に関する実験

レビューの信頼性評価における有効な属性と手法を特定するために、Amazon.com を対象として実験を行った。Amazon.com における 4 つのカテゴリ（ミステリー映画、ロックミュージック、MP3 プレーヤ、デジタルカメラ）に属する商品に対する全てのレビューとそのレビューを行ったレビューアのデータを独自に収集し、それをデータセットとした。

実験設計 これまでの研究から、レビューの信頼性評価に用いることができる方法は 2 つある。注目アイテムに対するレビューの内容をもとにレビューの質を評価し、レビューの評価値をレビューアの評価値

とみなす方法（以下、「内容に基づく手法」と記す）と、レビューアが過去にレビューを行った履歴をもとにレビューを直接評価する方法（以下、「履歴に基づく手法」と記す）である。前者の関連研究としては、[1, 2, 3] があり、後者の関連研究としては [4, 5] がある。我々は過去のレビュー内容を用いることによってレビューの書くレビューの質のバラつきを考慮できると考え、3 つ目の方法としてレビューアが過去に執筆したレビュー内容に基づく手法（以下、「過去の内容に基づく手法」と記す）を用いる。

本研究で用いる属性について述べる。内容に基づく手法と履歴に基づく手法の属性については、それぞれの手法の関連研究で用いられてきた属性を全て使用する。過去の内容に基づく手法の属性は内容に基づく手法の属性と同じものを用い、レビューアが過去に執筆したレビューにおける内容に基づく手法の属性値の平均を属性値とする。

有効な属性の特定は、各属性の属性値とレビューアの評価値の正解値との相関を計測することによって行う。レビューアの評価値の正解値には、Amazon.com での投票結果 (helpful と投票された割合) を用いる。有効な手法の特定は、それぞれの手法において使用する属性の属性値からレビューアの評価値を予測するモデルを学習し、学習したモデルによる予測値とレビューアの評価値の正解値との相関を計測することによって行う。モデルの学習にはサポートベクター回帰を用い、相関の計測には、ピアソンの積率相関係数を用いる。

実験結果 属性の有効性に関する実験の結果について述べる。4 つのカテゴリの全てのレビューにおける属性値とレビューアの評価値との相関係数の上位 10 属性を表 1 に示す。表 1 より、星の数に関する属性、商品特徴語に関する属性、単語数の相関が高くなっていることが分かる。

手法の有効性に関する実験の結果について述べる。信頼性評価の 3 手法に加え、3 手法を組み合わせたハイブリッド手法 (61 属性) (以下、「Hybrid61」と記

Copyright is held by the author(s).

* Yuya Tanaka, Nobuko Nakamura, Yoshinori Hijikata and Shogo Nishida, 大阪大学大学院基礎工学研究科

表 1. 正解値と各属性値との相関係数の上位 10 個

順位	属性名	相関係数
1	星の数 (過去)	0.5421
2	星の数	0.5359
3	平均星の数との差	0.4886
4	商品特徴語の数	0.3021
5	商品特徴語の数 (過去)	0.2799
6	単語数	0.2422
7	商品特徴語の頻度 (過去)	0.2292
8	商品特徴語の頻度	0.2197
9	単語数 (過去)	0.2034
10	被レビュー数	0.1880

す) と、過去の内容に基づく手法と履歴に基づく手法を組み合わせたハイブリッド手法 (33 属性)(以下, 「Hybrid33」と記す) も用いる. 各手法における予測値と正解値との相関係数の結果を表 2 に示す. 表 2 より, Hybrid33 が全てのカテゴリにおいて他の 4 手法よりも良い結果となっていることが分かる.

表 2. 各手法における予測値と正解値の相関係数

	内容	履歴	過去の内容	Hybrid61	Hybrid33
All	0.6165	0.5224	0.6282	0.6258	0.6419
Mystery	0.6505	0.5616	0.6826	0.6979	0.7216
Rock	0.7852	0.6552	0.7920	0.7902	0.7926
MP3	0.5595	0.4541	0.5680	0.5670	0.6138 *
DCamera	0.6389	0.5224	0.6452	0.6436	0.6524

“*”は他手法に対して統計的に有意であることを示す.
(相関係数の同等性の検定: 有意水準 5%)

3 システム設計

2章の実験において, レビューの信頼性評価において最も有効な手法は Hybrid33 であること, レビューの信頼性評価に対して有効な属性は, 星の数, 商品特徴語, レビューの分量に関する属性群であることを特定した. そこで, Hybrid33 を用いてレビューのランク付けを行い, 上記 3 つの属性群のスコアを提示することによりレビューの信頼性の説明付けを行うシステムを開発する. 星の数に関する属性群として “星の数 (過去)” と “平均星の数との差” を, 商品特徴語に関する属性群として “商品特徴語の数 (過去)” と “商品特徴語の頻度 (過去)” を, レビューの分量に関する属性群として “単語数 (過去)” と “文の数 (過去)” を用いてスコアを算出する. スコアの算出は, それぞれの属性の属性値に属性の有効性実験で得られた相関係数を重みとして掛け, それらを足し合わせることで行う.

上述のシステムを Web アプリケーションとして実装した. ユーザは, Web ブラウザから上記 Web アプリケーションにアクセスする. サーバは Java, クライアントは HTML で実装している. サーバは, Amazon.com からダウンロードしたレビューを基に, オフラインでレビューのスコアの算出を行う. ユーザがクライアントに商品ページの閲覧を要求す

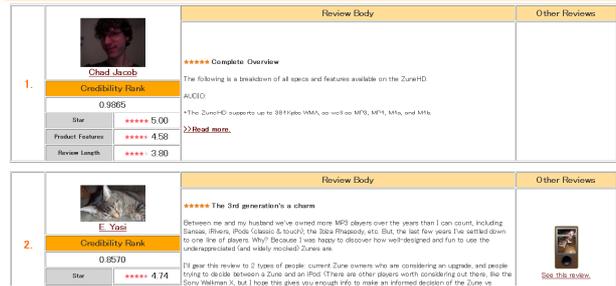


図 1. システムのスクリーンショット

ると, クライアントは要求された商品ページを開発したサーバからダウンロードする. さらに, ユーザが商品ページに埋め込まれたボタンをクリックすると, クライアントはその商品にレビューしたレビューのランキングを表示するページをダウンロードする.

図 1 にシステムのランキングページのスクリーンショットを示す. レビューランキングのページでは, レビューの信頼性のスコア (“Credibility Rank”), 信頼性に関する属性群のスコア, レビュー本文, 他の商品に対するレビューを見ることができる. ユーザは, 信頼性と関連する属性のスコアを見ることにより, そのレビューのどのような部分が評価されているのかを確認することができ, そのレビューが執筆したレビューを参考にするかどうかの判断材料とすることができる.

4 おわりに

本研究では, 実際の商用 Web サイトを用いてレビューの信頼性評価における有効な属性と手法を明らかにした. また, レビューのランク付けを行うとともにレビューの信頼性に関する属性の評価値も提示するシステムを開発した.

参考文献

- [1] S. Kim, et al., Automatically Assessing Review Helpfulness, *Proc. EMNLP'06, ACL*, pp. 423-430, (2006).
- [2] Z. Zhang, et al., Utility scoring of product reviews, *Proc. CIKM'06, ACM*, pp. 51-57, (2006).
- [3] J. Liu, et al., Low-Quality Product Review Detection in Opinion Summarization, *Proc. EMNLP-CoNLL'07, ACL*, pp. 334-342, (2007).
- [4] T. Riggs, et al., An Algorithm for Automated Rating of Reviewers, *Proc. JCDL'01, ACM*, pp. 381-387, (2001).
- [5] M. Chen, et al., Computing and using reputations for internet ratings, *Proc. EC'01, ACM*, pp. 154-162, (2001).