

中文スパムレビュー検出のための ショッピングサイトレビュー分析

Shopping Site Review Analysis for the Chinese Spam Review Detection

劉 真真 梶井 文人 プタシンスキ ミハウ

Shinshin Ryu Fumito Masui Michal Ptaszynski

北見工業大学 情報システム工学専攻

Kitami Institute of Technology, Dept. of Computer Science

There is a large amount of reviews on shopping sites, referring to which is becoming common when making a purchase. However, recently there have been noticed a problem of uploading fake or spam reviews which praise or depreciate the seller's goods. Particularly in China, this problem is noticeable, and the reliability of reviews is low. In this study, in order to automatically detect Chinese spam reviews, we analysed the review sentences in one of popular Chinese shopping site.

1 はじめに

ブログや SNS の普及に伴い、個人の意見や主張などを容易に発信する環境が整ってきた。これに伴い、アマゾン (www.amazon.co.jp) や楽天市場 (www.rakuten.co.jp) などのショッピングサイトにおいて大量のレビューが蓄積されており、ユーザは購買行動を行うにあたり、これらのレビューを参照することが常態化しつつある。

レビューの中には、購買行動を恣意的に誘導しようとするステルスマーケティングに類するものや、誹謗中傷などを目的としたもの、レビュー対象とは無関係のもの、商品広告を目的としたものや、他社がライバル社の商品の印象を悪化させる目的で書かれたスパムレビューが少なからず含まれており、サイト閲覧者数の減少やレビュー信頼性の低下などの悪影響をもたらしている。

その中、中国においては、人口を背景としたユーザ数の急拡大とともにスパムレビューの量も大規模化している。近年では「水軍」と呼ばれるプロのステルスマーケティングレビュアーが出現するなど、スパムレビューに対する早急な対応が求められている [1]。フェイクレビューは、オンラインショップの店長が自商品を褒めるや、他商品を貶めるために、「水軍」に大量な偽のレビューを頼むことで、スパムレビューの一種である。

しかし、中国語特有の言語構造や曖昧性、多様なネットスラングの影響もあり、中国語フェイクレビューの分析や研究のハードルは高く、問題点も多く存在している。

本研究では、中国語を対象としたフェイクレビュー検出技術の提案を目指す。特に、ステルスマーケティングの検出に注目し、中国語の言語特性の利用、レビュー解析によるユーザ心理の推定、製品やレビューに対する評判情報の解析などを組み合わせることを目指す。

以下、2章で従来の手法について紹介、3章でデータセット収集について説明、4章で中国ショッピングサイトのレビュー分析について述べ、考察する。

2 従来の手法について

フェイクレビュー検出は、ノーマルレビューとフェイクレビューの混合文書集合からフェイクレビュー集合を区別するという、一種の文書分類問題と考えることができる。そのため、フェイクレビュー特有の言語情報や記述者 (スパマー) の行動特徴を素性として教師有り学習による分類モデルを構築しようとする研究が多い [2]。

例えば、Ottら [3] は Amazon.com に投稿されたレビュー文書を収集し、書き手の文書スタイルの分析と心理学的手法を組み合わせることによってフェイクレビュー検出を試みている。Song [4] は、文法的特徴からフェイクレビューを分析している。李ら [5] は、レビュー文書の特徴に加えてアカウントや商品の特徴情報も利用することでより性能の高いフェイクレビュー検出手法を提案している。Fornaciariら [6] は、犯罪小説家の協力を得てフェイクレビューの特徴を含むデータセットを作成し、さらに人間判断をもとに学習したアルゴリズムによってラベル付けをすることでフェイクレビューの検出性能が向上したことを示している。

教師有り学習では、訓練データの作成において精密なラベル付けが必要となるが、フェイクレビュー検出においては、文脈やコンテンツ、トピック毎にフェイクであるか否かの判断基準が揺れるため、ラベル付け基準の標準化は非常に難しい。こうした問題を回避するために、教師無し学習を導入した研究も提案されはじめたが、高い有効性が確認されるには至っていない [7]。

上記に加えて中文レビューに関しては、語意の分析や自然言語の理解など解決すべき課題が多い [8]。特に、実験用の公開中文データが少ないこと、訓練データの精確性と信頼性が低いこと、中国のショッピングサイトにおけるユーザ公開情報が少なく、購買者の行為特徴が把握しにくいこと、中国語が持つ言語構造の特殊性や中国ネット社会における伝達方法の独自性などが挙げられる。

本研究では、中国語ショッピングサイトのレビューデータからフェイクレビューを検出するシステムの構築を目指す。今回は、中国語ショッピングサイトに記述されたレビュー分

析と、フェイクレビューの特徴についての考察を重視する。

3 データセット収集

分析対象として淘宝网 (world.taobao.com) に登録された衣料製品のレビューを用いた。

同サイトには 8 億以上の製品が並んでおり、約 5 億のユーザアカウントが存在し、平均的に 1 分で 4.8 万件の商品が売られているという中国国内最大手のショッピングサイトである [9]。2015 年 11 月 24 日現在では、ランキングトップの五つの衣料製品のレビュー数は 2 万件から 15 万件に及んでいる。

同サイトの販売ランキングに基づき、人気が高く、デザインが異なる 4 つの衣料製品を選び、それぞれの製品について書かれたレビューを各 50 件人手で抽出した。抽出した 4 製品について人気度ランキングに基づいて降順にソートした。今回はまずその第 1 位の製品を分析対象とし、ラベリングを行った。

レビューを参考にしたオンラインショッピングの経験を持つ 30 名の中国人を評価者として用いた。一つのレビューに対して、30 人の評価結果を統計し、半数以上がフェイクだと判断した場合はフェイクレビュー、そうではない場合はノーマルレビューとした。その結果、50 件中、フェイクレビューは 23 件であり、46%の割合となった。

以上の 30 名の中国人をグループ A とし、さらに 30 名の中国人を評価者グループ B とした。二つのグループの年齢、性別や、ショッピング頻度は同等になるように配慮した。

同じレビューに対し、A と B グループの判断する一致性を検証したところ、k 値 0.52 が得られた。フェイクレビュー判断の難しさを考慮すれば、一致性は十分だと考えられる。

4 中国ショッピングサイトのレビュー分析と考察

サイト内で設定されている評価項目を参考にしながら、14 個の製品評価パラメータ (デザイン、見た目、品質、《説明文との》一致性、サイズ、仕立て、生地、色落ち、《衣料製品を着たときの周囲からの》評判、値段、スタッフ、速達、総合満足度、勧め度) を設定し、パラメーターごとに人手でスコアリングした。

スコアは 1 から 5 までの 5 段階の評定値 (とても不満、不満、どちらでもない、満足、とても満足) で表現した。図 1 に、分析対象とした衣料製品のスコアリング結果の度数の例を示す。横軸に評定スコア、縦軸にはそれぞれのパラメータを含むレビュー数を示している。

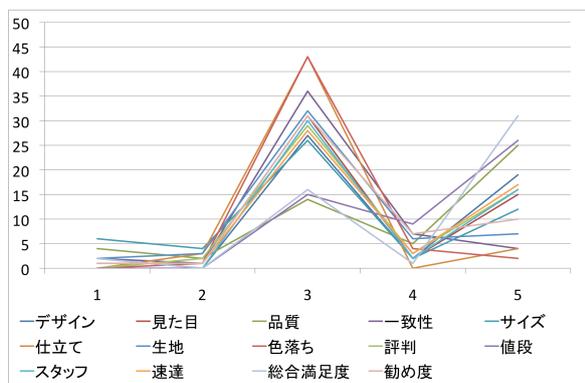


図 1: 衣料製品のスコアリング結果の度数分布

図 1 では、スコア 3 のレビューの数が最も多い。これはレビュー中にパラメータに関する記述がない場合、スコア 3 「どちらでもない」と判断するからである。つまり、全

面的に評価するレビューの割合は少ない。特に「色落ち」、「仕立て」と「一致性」のスコア 3 が非常に多い。

「値段」、「品質」と「総合満足度」という三つパラメーターにおいては、スコア 5 が一番多い。その原因を調べると、対象とした製品は約 400 円であるため、安いと判断され、「値段」にとっても満足する人が多かった。「品質」に関しては、スコア 5 の (褒める) レビューが非常に多いことがわかった。また、ノーマルレビューの内容を調べると、品質面において満足するユーザも存在している。「総合満足度」のスコア 5 も比較的が多いことから、値段が非常に安く、満足する人が多いと考えられる。

「総合満足度」はユーザが今回の購買に対する満足度を表すパラメーターである。しかし、複雑性と曖昧性が高く、スコアが正確に反映されていないため、今回は「総合満足度」を考慮しないことにした。

次に、パラメーターごとのスコアの頻度分布に基づいた分析を試みる。95%の信頼区間を計算し、これを用いて分析を行った。信頼区間外、すなわち外れ値は何らかのバイアスが掛かっている可能性があり、主なバイアスとして本質的な製品への言及を逸脱した、意図的な記述が考えられる。このような記述の多くはフェイクレビューである可能性が高いと考えられる。パラメーターごとに信頼区間以外のフェイクレビューの割合は表 1 のようになる。

(計算方法については、「信頼区間以外のフェイクレビューの割合 = 信頼区間以外のフェイクレビューの数 / 信頼区間以外の全部のレビューの数」である。)

表 1: 信頼区間以外のフェイクレビューの割合

パラメーター	割合
勧め度	0.83
スタッフ	0.61
評判	0.61
仕立て	0.57
見た目	0.53
デザイン	0.52
一致性	0.50
速達	0.47
色落ち	0.43
生地	0.42
サイズ	0.39
値段	0.20
品質	0.00

(表 1 には割合高い順で示し、割合が高いほど、フェイクレビューの特徴も明らかだと考え、高い値と低い値のパラメータをそれぞれ四つ分析を行った。)

「勧め度」、「スタッフ」、「評判」と「仕立て」は外れ値よりフェイクレビューの可能性が高いものが多く、「生地」、「サイズ」、「値段」と「品質」のその割合が少なかった。そして、上位 4 件と下位 4 件について、より詳しく議論を行う。

上位の「勧め度」、「スタッフ」、「評判」と「仕立て」のスコアの内訳を図 2 と図 3 に示す。図 2 はフェイクレビュー、図 3 はノーマルレビューである。ここで、「どちらでもない」のスコア 3 の結果を除いて、主に意見が含めた結果を分析した。

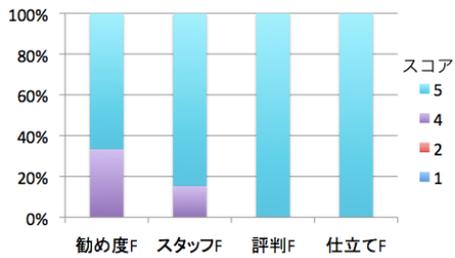


図 2: フェイクレビューに対するスコア内訳 (外れ値割合上位)

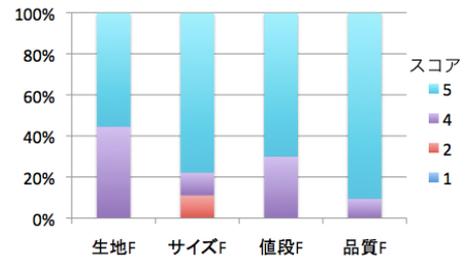


図 4: フェイクレビューに対するスコア内訳 (外れ値割合下位)

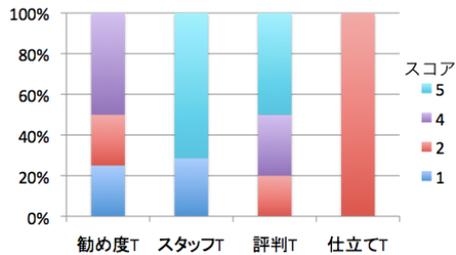


図 3: ノーマルレビューに対するスコア内訳 (外れ値割合上位)

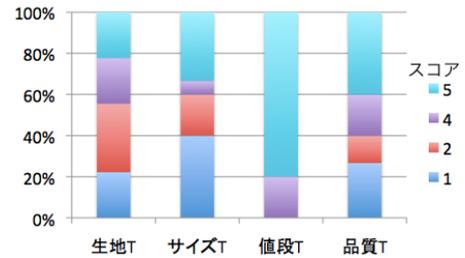


図 5: ノーマルレビューに対するスコア内訳 (外れ値割合下位)

全体的な傾向として、フェイクレビューの方はスコア5が多く、ノーマルレビューの方は低いスコアがより多かった。それはフェイクレビューが褒める内容が多く、ノーマルレビューに様々な意見が含まれているためと考えられる。

ノーマルとフェイクレビューの内訳には差が見られた。特に、「勧め度」、「評判」と「仕立て」の差が大きく、具体的にみると、「勧め度」には、フェイク内訳はスコア4と5のみでお目られており、強く勧める記述ばかりである。「評判」についても、フェイク内訳は全てが褒めるコメントであり、ノーマル内訳は様々な意見が含まれている。「仕立て」については、フェイク内訳は全てがスコア5、ノーマル内訳は全て2であった。このことから、多くのユーザーは「仕立て」に対して不満を持っているのに、極端に褒めるフェイクレビューが混在していることを示している。

下位のパラメーターである「生地」、「サイズ」、「値段」と「品質」のスコア内訳を図4と図5に示す。全体的には、フェイクレビューはスコア5が多い、ノーマルレビュー内訳との明確な差異が見られなかった。それはフェイクレビューの割合が低く、フェイクの特徴が表面化しなかったためと考えられる。

その中では「生地」、「サイズ」と「品質」においてより大きな偏りが見られた。具体的には、「生地」のフェイクレビューが褒めるレビューが多いのに対して、ノーマルレビューは様々なコメントが含まれていた。「サイズ」のノーマルレビューではスコア1の不満な人が多かった。「品質」の内訳を見ると、フェイクレビューの方が褒めるレビューが多いが、ノーマルレビューでは様々な意見があった。

5 おわりに

本稿では、ショッピングサイト上に激増しているフェイクレビューを検出するために、ショッピングサイトのレビューを分析した。今回は対象とした第1位の衣料製品のレビューをラベリングし、14個のパラメーターを人手でスコアリングし、分析を行った。フェイクレビューは褒める内容が多いことがわかった。フェイクレビューの特徴が明らかなパラメーターを用いて、それらのキーワードをフェイクパターンとして収集し、フェイクレビューの分類器に使う予定である。

次回からの評価はA、Bグループで評価する予定である。一つの製品のレビューは50件あり、25件をグループAで評価し、残りの25件をグループBで評価する。

これから収集したパターンを用いて、フェイクレビューを検出するシステムを作成する。今の段階で、本研究で用意したラベル付きデータを用いて、それを訓練データとし、Ptaszynskiら[10]のSPECシステムを使って、フェイクレビューの自動分類実験を行い、その結果を探討する予定である。

参考文献

- [1] Cheng Chen, Kui Wu, Venkatesh Srinivasan, Xudong Zhang, “Battling the internet water army: detection of hidden paid posters,” Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, pp.116-120, New York, NY, USA.
- [2] 趙衍, “ネットワークレビュー研究述評”, 『上海管理科学』, (2014.8).
- [3] Myle Ott, Yejin Choi, Claire Cardie, Jeffrey T. Hancock, “Finding Deceptive Opinion Spam by Any Stretch of the Imagination,” Proceedings of the 49th Annual Meeting of the ACL, pp.309-319, Portland, Oregon, June 19-24, 2011.
- [4] Song Feng, Longfei Xing, Anupam Gogar, Yejin Choi, “Distributional Footprints of Deceptive Product Reviews,” Association for the Advancement of Artificial Intelligence (www.aaai.org), 2012.
- [5] 李霄, 丁晟春, “ゴミ商品レビューの識別研究”, 『現代図書情報技術』, pp.63-68(2013.1).
- [6] Tommaso Fornaciari, Massimo Poesio, “Identifying fake Amazon reviews as learning from crowds,” Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics, pp. 279–287, Gothenburg, Sweden, April 26-30, 2014.
- [7] 宋海霞, 嚴馨ら, “半教師あり学習を用いたフェイクレビューの検測”, 『昆明理工大学学報』, pp.59-65(2015.10).
- [8] 許少岩, 鍾敏娟, “ネット商品レビュー中のゴミレビューの識別方法について”, 『科技広場』, pp.231-238(2014.2).
- [9] www.taobao.com/about/intro.php?spm=0.0.0.0.P1tDEG
- [10] Michal Ptaszynski, Rafal Rzepka, Kenji Araki, Yoshio Momouchi, “Language Combinatorics: A Sentence Pattern Extraction Architecture Based on Combinatorial Explosion,” International Journal of Computational Linguistics (IJCL), Volume (2) : Issue (1) : pp. 24–36, 2011.