

電化製品レビュー抽出のための Blog 記事内容分析

Content Analysis of Blogs for Extracting Electric Appliances Reviews

渡邊 裕介^{*1}
Yusuke Watanabe

松尾 朋子^{*1}
Tomoko Matsuo

安藤 一秋^{*2}
Kazuaki Ando

^{*1} 香川大学大学院工学研究科
Graduate school of Engineering, Kagawa University

^{*2} 香川大学工学部
Faculty of Engineering, Kagawa University

In recent years, review sites are utilized as information sources for decision making of product purchases. However, the number of reviews in review sites includes deflection. Some of them have no or few reviews to be referred. Therefore, the purpose of our study is to extract reviews from blog articles automatically. This paper describes content analysis of blog articles for extracting electric appliances reviews as a preliminary step.

1. はじめに

Blog や電子掲示板, 口コミサイトなどユーザが情報を発信する CGM (Consumer Generated Media) サイトの利用者が増加し続けている。その中でもレビューサイトは, 商品購入の意思決定時に情報源として頻繁に利用される[安藤 12, 石橋 11, 乾 06, 落合 11, 小林 05, 河野 08, 佐々木 11, 杉木 08, 関根 10, 立石 01, 田中 11, 平山 12, 渡邊 13]。しかし, レビューサイトにおける商品毎のレビュー数には偏りがあり, レビューが書かれていない, レビュー数が少ない商品も数多く存在する。また, レビューが書かれていても, 自分が必要とする情報が書かれていない場合は参考にならない。したがって, レビューサイト以外の情報源から有用なレビューを抽出し, 情報量を補う必要がある。

本研究では, レビューの情報量を補うメディアとして, Blog に着目する。Blog には, 個人の経験, 感想などが記載されているため, 有用なレビュー情報が書かれている可能性が高い。しかし, Blog は構造化されていないため, 検索エンジンを利用しても有用なレビューを横断的に閲覧できない。そこで本研究では, Blog からレビューを自動抽出することを目的とする。本稿では, その前段階として, レビューサイトに記述される電化製品のレビュー数と Blog 記事数の関係について分析する。

2. 研究対象とするレビューの定義

Amazon, 楽天, 価格.com からレビュー数の多い任意の電化製品レビューを各々約 100 件調査した結果, 商品を実際に使用した経験が書かれているレビュー, 様々な項目に対して詳しく評価しているレビューは, レビューアの評価が高い傾向があることが分かった。

そこで, 本研究で抽出対象とするレビューを「レビューアが購入し, レビューアが使用した経験情報」と定義する。そのため, 商品を購入しただけで経験情報が記載されていない情報は, レビューに含まない。また, 「店頭で使ってみた」などの商品を購入していない一時的な経験情報や, 「友人が使用してよかったと言っていた」のようなレビューアの経験では無い情報もレビューに含まない。

3. レビューサイトの電化製品(商品)とレビュー数の関係調査

レビューサイトにおける商品とレビュー数の関係を調査するた

連絡先: 渡邊裕介, 香川大学大学院工学研究科, 高松市林町
2217-20, s12g485@stmail.eng.kagawa-u.ac.jp

めに, 価格.com における家電カテゴリー全商品のレビュー数毎の商品分布を調査する。調査には, 商品の一覧ページから機械的に抽出したレビュー数を用いる。この時, 色は商品の性能に関係なく, 個人の好みにより決定するもの, またレビューページが同じ情報が提示されることから, 色違いの商品について製品型番の集約を機械的に行う(例: NW-F807(W), NW-F807(B) → NW-F807)。その結果, 対象商品数は 52,198 商品となった。

調査の結果, 全 52,198 商品中 40,268 商品(77.1%)にはレビューが存在しなかった。レビューが存在しなかった商品例を表 1 に示す。表 1 に示すように, レビューが存在しなかった商品には, 冷蔵庫・冷凍庫や洗濯機のように頻繁に買い替えない大型家電やカセットテープのように使用されることが少なくなった商品が多かった。

また, 価格.com に存在するレビュー数は全部で 90,173 件であった。そのため, 1 商品の平均レビュー数は, 1.7 件(90,173 件/52,198 商品)となる。

以上の結果より, 幅広いユーザを対象とする場合, レビュー数の少ない商品に対し, レビューを補う必要があると考えられる。

表 1: レビューが存在しなかった商品例

商品カテゴリー	メーカー	規格名
液晶テレビ	日立	REGZA 32H1000
ブルーレイ・DVDレコーダー	SONY	スゴ録 RDR-GX5
洗濯機	シャープ	イオンコート ES-FG55F
冷蔵庫・冷凍庫	ナショナル	NR-A51T
カセットテープ	マクセル	DAT DM-120D

4. レビューサイトの商品レビュー数と Blog 記事数の関係調査

レビューサイトの商品レビュー数と Blog 記事数の関係の有無を調査する。調査には, 3. で利用した価格.com から抽出したレビュー数と Yahoo! Blog 検索 Web API で取得した Blog 記事数を利用する。なお, Blog の検索キーは, 商品の一覧ページから機械的に抽出した 52,198 個の製品型番を用いる。

調査の結果, レビュー数と Blog 記事数が 0 件の商品が全 52,198 商品中 35,099 商品(67.2%)と最も多いことがわかった。また, Blog 記事数が増加するに従って, 商品数が減少することがわかった。また, レビュー数が 0 件の商品 40,268 商品中 35,099 商品(87.2%)は, Blog 記事が存在しなかった。

以上の結果をまとめる。製品型番を含む Blog 記事数を利用した調査ではあるが、Blog 記事数が 0 件の商品が多いことが確認できた。したがって、多くの商品は、Blog を用いても必ずレビューを補えるとは限らないと考えられる。

5. レビュー数と Blog 記事数の商品種別における関係調査

3. と 4. の調査で利用したデータを利用し、レビュー数または Blog 記事数の多い、もしくは少ない商品の特徴について、商品の種別単位(液晶テレビ、冷蔵庫など)で調査する。ただし、本調査に利用する Blog は製品型番で検索したものであるため、すべての Blog にレビューが書かれているとは限らない。

調査の結果、全商品種 114 種中 95 種(83.3%)は、レビュー数より Blog 記事数の多いことがわかった。レビュー数に比べ、特に Blog 記事数が多い商品種には、興味があるが高価であるため購入者が少ないと考えられるもの(例:放射線測定器など)、熱狂的なファンがいると考えられるもの(例:アンプ、スピーカーなど)などが多く見られた。

これらの商品に対する Blog の内容を調査した結果、前者は、実際に購入することが困難であると考えられるため、レビューではなく「欲しい」などと言った願望が書かれていた。後者は、こだわりのあるユーザが自身の趣味について日常的な話題の一つとして記述したことで、Blog 記事数が多くなったと予想される。

また、全 114 種中 14 種(12.3%)は、レビューが 0 件であった。その 14 種中 12 種(85.7%)は、Blog 記事の存在が確認できたが、12 種中 11 種(91.7%)は、消耗品(例:レコード針、フィルターなど)であった。これらの品種に対して Blog からレビューを補っても、利用するユーザ数は少ないと考えられる。

Blog 記事数よりレビュー数が多い商品は、全 114 種中 19 種(16.7%)あった。これらの商品の傾向としては、日常的に使用される品種(例:テレビ、冷蔵庫など)が多く見られた。幅広いユーザ層が日常的に使用するため、レビューが書きやすいと予想される。

以上の結果より、すべての商品についてレビューを補うことは難しいが、特定の商品に限定することで、レビューを補える可能性があるといえる。

6. Blog 記事の内容分析と考察

6.1 調査

商品の型番で収集した Blog 記事群にレビューが含まれる割合を調査するため、Blog の内容を調査する。本調査では、Blog 記事の内容を、「レビュー」、「広告」、「商品違い」、「抽出ミス」、「その他」に分ける。また、「その他」の分類に、「商品名のみ」、「レビュアー以外の経験情報」、「非購入レビュアーの経験情報」、「対象商品と他商品の商品情報比較」、「対象商品と他商品の比較」、「対象商品の情報まとめ」、「商品名のみ」に分類して考察する。

調査対象は、Blog 記事数とレビュー数の関係を考慮し、以下の 6 種類とする。なお、各商品種カテゴリーの調査結果より、Blog 数対レビュー数比は、3:1 が最も多かったことから、⑤と⑥を対象に加えた。

- | | | |
|---|----------------------|------------|
| ① | Blog 記事数 >> レビュー数 | : レビュー数多い |
| ② | Blog 記事数 >> レビュー数 | : レビュー数少ない |
| ③ | Blog 記事数 << レビュー数 | : レビュー数多い |
| ④ | Blog 記事数 << レビュー数 | : レビュー数少ない |
| ⑤ | Blog 記事数 ÷ 3 ≒ レビュー数 | : レビュー数多い |
| ⑥ | Blog 記事数 ÷ 3 ≒ レビュー数 | : レビュー数少ない |

調査対象とする各カテゴリーは、「①スピーカー」、「②放射線測定機」、「③液晶テレビ」、「④生ごみ処理機」、「⑤掃除機」、「⑥電球・蛍光灯」から任意で抽出した Blog 記事各 50 件を用いる。なお、「④生ごみ処理機」と「⑤掃除機」は、50 件取得できなかったため取得できた最大数とする。

表 2 に分析結果、表 3 にその他の詳細を示す。以下、表を基に、各商品について考察する。

(1) スピーカー

表 2 に示すように、最も多かった記事は抽出ミスの記事であり、50 記事中 20 記事(40.0%)存在した。抽出ミスの記事が多くなった理由として、型番の集約時に型番が短くなりすぎたため、その商品が記述されていない記事(例:テレビの型番やルータ名と一致)がヒットしたと考えられる。また、「その他」は、50 記事中 17 記事(34.0%)存在し、二番目に多い結果となった。この分類には、「スピーカーの聞き比べをしました。」のように実際にスピーカーを購入していないが、使用感を記述した Blog 記事が多かった。レビュー記事は、50 記事中 5 記事(10.0%)しかなかった。スピーカーのレビューを記述していたユーザは、記述内容からスピーカーのマニアであると考えられ、他のスピーカーとの比較内容を記述していた。また、50 記事中 4 記事(8.0%)は広告であった。広告記事には、商品の URL や商品が購入できる Web サイトの URL が記述されていた。

表 3 に示すように、「その他」の 17 記事中 16 記事(94.1%)には、商品の名前のみが書かれていた。例えば、「スピーカー〇〇とオーディオ××を組み合わせると・・・」といったように組み合わせの例示として記述されていた。17 記事中 1 記事(5.9%)は、他商品との比較結果を書いた記事であった。しかし、対象商品についての内容は記述されておらず、他商品との比較のみが書かれていた。

(2) 液晶テレビ

表 2 に示すように、最も多かった記事は、「その他」であり 50 記事中 23 記事(46.0%)存在した。これらの記事には、実際に購入して使用しているが使用した経験情報ではなく、「〇〇のテレビ使っている」というような報告が多かった。また、液晶テレビを購入する際、「〇〇と××で迷った」などのように比較候補に出されていることが多かった。次に多かった記事は、レビュー記事であり、50 記事中 14 記事(28.0%)あった。レビュー記事には、液晶テレビを購入した理由などがしっかり書かれている記事や「〇〇が悪かった」などのように使用して悪かった点を挙げている記事、液晶テレビが故障した内容が書かれた記事があった。なお、50 記事中 13 記事(26.0%)は、広告であった。広告記事には、スピーカーと同様、商品の URL や購入できるサイトの URL が記述されていることが多かった。

表 3 に示すように、「その他」の 23 記事中 15 記事(65.2%)は、商品名しか記述されていない記事であった。商品名しか記述されていない記事は、「〇〇を買った」や「〇〇を使っている」のように報告する記事であった。23 記事中 3 記事(13.0%)は、購入していないが経験情報を記述した記事であった。この記事には、「友人が使っている」や「家族に買ってあげてどうだった」のように他人からの経験情報を書いた記事であった。23 記事中 2 記事(8.7%)は、購入していないが使用感を書いた記事であった。この記事には、試聴品のテレビを使用した感想が書かれていた。23 記事中 2 記事(8.7%)は、商品比較情報が書かれた記事であった。商品比較情報が書かれていた記事には、どんな商品があるのかをまとめた記事やどの商品と迷ったかを書いた記事があった。23 記事中 1 記事(4.3%)には、商品を購入した理由が書かれていた。

表 2: Blog 記事の内容分析結果

	レビュー	広告	商品違い	抽出ミス	その他	合計	分類	レビュー数
スピーカー	5	4	4	20	17	50	①	多い
液晶テレビ	14	13	0	0	23	50	②	多い
放射能測定機	7	12	9	0	22	50	③	少ない
生ごみ処理機	8	1	0	0	7	16	④	少ない
電球・蛍光灯	7	17	0	0	17	41	⑤	少ない
掃除機	7	19	4	0	20	50	⑥	多い
合計	48	66	17	20	106	257		

表 3: 「その他」商品について詳細

	商品名のみ	レビュアー以外の 経験情報	非購入 レビュアーの 経験情報	対象商品と 他商品の商品 情報比較	対象商品と 他商品の比較	対象商品の情 報まとめ	合計
スピーカー	16	1	0	0	0	0	17
液晶テレビ	15	3	2	1	1	1	23
放射能測定機	21	1	0	0	0	0	22
生ごみ処理機	3	2	0	0	0	2	7
電球・蛍光灯	10	0	0	0	1	6	17
掃除機	8	0	1	3	0	8	20
合計	78	7	3	4	2	17	111

(3) 放射能測定器

表 2 に示すように、最も多かった記事は、その他の記事であり、50 記事中 22 記事(44.0%)存在した。これらの記事には、商品を使った感想ではなく、商品を使った測定結果について記述したものが多かった。続いて、広告の記事が多く、50 記事中 12 記事(24.0%)存在した。商品違いの記事は、50 記事中 9 記事(18.0%)存在した。この原因としては、型番が短いもの、特に数字のみのものが誤検索の原因であった。最後に、レビュー記事は、50 記事中 7 記事(14.0%)存在した。レビュー記事には、使用した感想、また、使用中の動画を記載しているものがあつた。

表 3 に示すように、「その他」の 22 記事中 21 件(95.5%)は、商品名のみが記述された記事であつた。商品名のみが記述された記事には、放射能測定機を使用して計測した値が書かれていた。22 記事中 1 記事(4.6%)は、他のユーザのレビューをまとめたものが記述されていた。

(4) 生ごみ処理機

表 2 に示すように、最も多かった記事はレビュー記事であり、16 記事中 8 記事(50.0%)存在した。また、広告記事は 16 記事中 1 記事(6.0%)存在した。広告記事が少ない要因として、生ごみ処理機が必要だと感じる人が少なく、広告に載せても買う人が少ないからではないかと考えられる。また、「その他」の記事が 16 記事中 7 記事(43.8%)存在し、これらの記事には、レビュアー以外の経験情報や購入を検討している人による商品についての比較や情報のまとめ記事など有用なものが多かった。しかし、家電製品販売の店長によるブログ記事があり、有用な経験情報と判断してよいかどうかを考えるべき情報もあつた。

表 3 に示すように、「その他」の 7 記事中 3 記事(42.9%)は、商品を購入した報告であり、商品名のみ記述されていた。7 記事中 2 記事(28.6%)は、他のユーザのレビューを記述している

記事であつた。また、7 記事中 2 記事(28.6%)は、商品の仕様情報について詳しく記述されていた。

(5) 電球・蛍光灯

表 2 に示すように、最も多かった記事は、「その他」と「広告」であり、41 記事中 17 記事(41.5%)存在した。「その他」の記事には、「この電球・蛍光灯を使用しています」や「取り替えました」などの記載が多かった。広告記事には、電球・蛍光灯の URL が記述されていた。そして、レビュー記事は、41 記事中 7 記事(17.0%)で、電球・蛍光灯に交換して明るくなったなどの経験情報が書かれていた。商品違いの記事は存在しなかつた。

表 3 に示すように、「その他」の 17 記事中 10 記事(58.8%)は、使っている電球の紹介や購入した報告であり、名前のみが記述された記事であつた。17 記事中 6 記事(35.3%)は、購入した理由や対象商品の説明を記述した商品まとめ記事であつた。17 記事中 1 記事(5.9%)は電球・蛍光灯の長さ比較を行った商品比較記事であつた。

(6) 掃除機

表 2 に示すように、最も多かった記事は、「その他」であり、50 記事中 20 記事(40.0%)存在した。これらの記事には、「〇〇を買おうか検討している」といった記事が多かった。次に多かった記事は広告記事であり、50 記事中 19 記事(38.0%)存在した。広告記事には、商品の紹介や商品の URL、購入できるサイトの URL が記述されていた。そして、レビュー記事は、50 記事中 7 記事(14.0%)であつた。レビュー記事には、掃除機を購入した理由や組み立て方法、使用した感想が記述されていた。最後に、50 記事中 4 記事(8.0%)は、商品違いの記事であつた。商品違いの記事では、おもちゃの名前の一部が掃除機の型番と一致していた。

表 3 に示すように、「その他」で名前のみが紹介されていたのは、20 記事中 8 記事(40.0%)であり、「うちはこの掃除機使って

ます」といった報告であった。また、20 記事中 8 記事(40.0%)は対象商品について書かれた記事であった。これらの記事には、購入理由や商品情報・仕様などが書かれていた。20 記事中 3 記事(15.0%)には、他商品との比較が書かれていたが、「〇〇と××では音の大きさが違う」といった程度の内容であり、詳しくは記述されていなかった。20 記事中 1 記事(5.0%)には、他人が購入した商品を借りた情報が書かれていた。

6.2 考察

全体の結果について考察する。全 257 記事中 47 記事(18.3%)がレビュー記事、257 記事中 63 記事(24.5%)が広告記事だった。この結果より、レビューより広告の方が多く分かる。広告には典型的なパターンが存在することが確認できたので、多くの広告は容易に排除可能である。

本調査で利用した Blog は、型番で検索し収集したものである。一般のユーザがレビューを記述する際、型番ではなく、商品名の一部や抽象化した名前を利用することも考えられるため、レビューの収集漏れが存在する。したがって、レビュー数が少なくなった可能性がある。

Blog 記事数に比べ、レビュー数が多い生ごみ処理機と液晶テレビについては、それぞれ 50.0%と 28.0%の割合で Blog 記事がレビューを含んでいる。したがって、Blog 記事を利用することでレビューを補えると可能性が高い商品であると考えられる。

一方、レビュー数に比べ、Blog 数が多い放射能測定器とスピーカーについては、Blog 記事がレビューを含む割合がそれぞれ 14.0%と 10.0%となり、期待するほどレビューを補えない可能性も考えられる。

調査の中で、その他に分類した記事の中には、レビューではないが有用な情報も多く見られた。そのため、これらを抽出するためにさらに分析をする必要があると考えられる。

目的の商品ではないものが、257 記事中 37 記事(14.4%)含まれていた。これは、型番を集約しすぎたことが原因である。また、長い型番の商品は、Blog 記事を抽出できない場合が多かった。そのため、検索語として型番以外の情報を用いる方法を考える必要がある。

7. おわりに

本稿では、電化製品レビューサイトに記述されるレビュー数と Blog 記事数の関係について分析した。分析の結果、レビュー数と Blog 記事数が共に 0 件の商品が最も多いことがわかった。また、特定の商品に対しては、Blog からレビューを補える可能性があることがわかった。

Blog 記事の内容調査を行った結果、Blog 記事数に比べ、レビュー数が多い商品が最もレビューを補える可能性が高いことがわかった。また、レビューとはいえないが有用である情報が多く存在することを確認した。したがって、それらの情報を抽出するための分析を行う必要がある。検索語に用いた型番については、短いもの、長いものに対する扱いについて考慮する必要がある。

今後は、その他に分類した Blog 記事内容について調査を行い、レビュー以外の有用な経験情報の抽出・判定手法について検討する。

参考文献

[安藤 12] 安藤 まや, 石崎 俊:インパクトの視点に基づく WEB 上のユーザレビューの分析, 言語処理学会年次大会第 18 回年次大会論文集, pp.731-734, 2012.

[石橋 11] 石橋 直己, 乾 孝司, 山本 幹雄:商品に関する被参照性と類似性に基づくレビュー文書からの競合商品同定手法, 言語処理学会年次大会第 18 回年次大会論文集, pp.863-866, 2011.

[乾 06] 乾 孝司, 奥村 学:テキストを対象とした評価情報の分析に関する研究動向, 自然言語処理, Vol.13, No.3, pp.201-241, 2006.

[落合 11] 落合 恵理香, 小林 一郎:商品の評価を対象としたレビュー文書の分析, 言語処理学会年次大会第 18 回年次大会論文集, pp.1176-1179, 2011.

[小林 05] 小林のぞみ, 乾健太郎, 松本裕治, 立石健二, 福島俊一:意見抽出のための評価表現の収集, 自然言語処理, Vol.12, No.2, pp.203-222, 2005.

[河野 08] 河野 洋志, 柴田 知秀, 黒橋 禎夫:ブログ記事の商品カテゴリへの自動マッピング, 言語処理学会年次大会第 14 回年次大会論文集, pp.733-736, 2008.

[佐々木 11] 佐々木 稔, 新納 浩幸:商品タイトルから商品名を自動抽出するための効率的な教師データ作成手法, 言語処理学会年次大会第 18 回年次大会論文集, pp.1015-1018, 2011.

[杉木 08] 杉木 健二, 松原 茂樹:消費者の意見に基づく商品検索, 情報処理学会論文誌, Vol.49, No7, pp.2598-2603, 2008.

[関根 10] 関根 聡, 小林 暁雄, 坂地 泰紀, 竹中 孝真:ショッピングサイトにおける商品の同一性・類似性の推定手法, 言語処理学会年次大会第 16 回年次大会論文集, pp.254-257, 2010.

[立石 01] 立石健二, 石黒義英, 福島俊一:インターネットからの評判情報検索, 情報処理学会研究報告, 2001-NL-144, Vol.2001, No.69, No.11, pp.75-82, 2001.

[田中 11] 田中 俊行, グェン ミンティ, 中川 博之, 田原 康之, 大須賀 昭彦:評判分析システムのための教師あり学習を用いた意見抽出, 電子情報通信学会論文誌, Vol.J94-D, No.11, pp.1751-1761, 2011.

[平山 12] 平山 拓央, 湯本 高行, 新居 学, 佐藤 邦弘:語の共起と極性に基づく商品レビュー閲覧支援システム, 情報処理学会研究報告, Vol.2012-DBS-155, No.3, pp.1-9, 2012.

[渡邊 13] 渡邊尚吾, 乾孝司, 山本幹雄:カテゴリ情報を利用した blog 記事からの商品名自動抽出, 言語処理学会年次大会第 19 回年次大会論文集, pp.3-6, 2013.