

評価対象と利用目的を考慮したレビュー文の意見分類

Classifying Opinions in Product Reviews to Fulfill Developer Demands.

林 光雄
Mitsuo Hayashi

大熊 智子
Tomoko Ohkuma

三浦 康秀
Yasuhide Miura

谷口 元樹
Motoki Taniguchi

富士ゼロックス(株) 研究技術開発本部 コミュニケーション技術研究所
Communication Technology Laboratory, Research & Technology Group, Fuji Xerox Co., Ltd.

In recent years, reviews about various products are posted on the web to share opinions about them. The reviews include opinions such as evaluations, requests, and bug reports that are valuable for the developers of the products. However, these opinions are often unorganized and are hard to analyze. In this paper, we define four opinion attributes of product reviews that can help developers to find information in certain purposes. The paper also shows that by using some machine learning techniques, we would be able to automatically classify reviews with these four attributes.

1. はじめに

1.1 背景

顧客の要望を早急に製品やサービスに反映することは、それを提供する企業または個人にとって重要な課題であり、その為にコールセンターやホームページ、メールなどを通じて顧客の声が集められている。最近では web 上の商品レビューにも企業側にとって有用な情報が、顧客にあたるユーザによって投稿されている。

商品レビューにはある製品やサービスを対象にその評価や意見が書かれる。本来のレビューを利用する目的は、主にその商品を未所有のユーザがその商品の購入・導入を決定するための参考情報とすることである。しかし最近のレビューサイトでは商品以外のものに対する意見(例えば他のレビュー投稿者への意見、サービス提供者の運営方法に対する意見など)も書きこまれるようになってきている。さらに、要望や不具合報告など商品提供側にとって有用な情報も含まれていることがある。このことから新たな利用目的として、商品を提供する企業側が不具合修正などの改善を速やかに行う為に利用することが考えられる。しかし前述の通り、対象の異なる評価も含まれるため、商品に対する評価や不具合情報をレビューから読み取ることは、従来の収集された顧客の声に比べると困難である。

1.2 本研究の目的

そこで、本研究は商品のレビュー文を対象として、その商品を提供している企業または個人が商品の修正や次回製品の開発を行う為に有用な情報を含む文を効率よく識別することを最終的な目的とする。そのために、本稿では 1)利用目的を考慮して分析しレビュー文コーパスを作成し、2)機械学習による自動分類手法を提案する。2)では人に対するものと商品に対するものを分離する為に、人称代名詞が出現したかどうかを示す素性を設定し、その効果の有無を確認した。

2. 関連研究

商品やサービスに対するレビュー投稿から評判情報を抽出する研究は従来から行われてきた。例えば、[Turney 2002]はレビューに記載されたテキストと評価値の関係を学習し、テキスト

から評価値を予測する研究を行っている。[Dave 2003]はレビューの内容と評価値の不整合や評価者の個人差を考慮して、テキスト処理による極性(ポジティブ/ネガティブ)の判定を行っている。[飯田 2005]はレビュー文から特定の対象(デジカメと車)に関する記事を抜き出し、それに対して評価対象、評価値、評価属性の三つ組情報を抽出することを試みている。さらに、[Stoyanov 2008]ではレビューに限らず、ネット上に投稿されたテキストを対象として、極性やその度合いだけではなく、評価している対象が何であるかまで分析の対象としている。[藤村 2004]はレビュー文ではなく掲示板を対象にした評判抽出を行っているが、評判情報抽出については既にできていると仮定した上で、その評判の極性判定を行っている。

本研究はレビューを対象にしているが、そこに含まれるテキストが必ずしもそのレビュー対象に対する評価ではないことを考慮するものである。また、ここで提案する手法は[Stoyanov 2008]や[飯田 2005]のように評価対象について細かく特定は行わないが、人に対するものと商品に対するものを分離することを特徴としている。

3. 意見属性の定義

3.1 方針

対象とするレビュー文はスマートフォン用アプリ配信サイトのゲームのレビューから抽出した。これはユーザ数が多く、表現の自由度が高いためである。次に、レビュー文の利用目的として下記の二つを仮定した。

1. 新規ユーザが購入や採用を決定するための参考意見として利用する
2. 商品提供側の開発者や運営者が現状を把握する為に利用する

この二つの観点からレビュー文の分析を実施し、レビュー文へ付与する4種の意見属性(報告, 要望, 評価, クレーム)を定義した。各属性をレビュー文の1文毎に付与する。1文に複数の属性が当てはまる場合は複数の属性を付与する。それぞれの属性について下記に詳しく述べる。

3.2 属性

(1) 報告

利用目的は、主に製品・サービスを提供する側の開発者などが自社製品の不具合等を素早く把握することである。[報告]に

は事象の内容と、それが起こった背景が記載されている必要がある。特に不具合が起こっている場合には、それを再現して詳細を確認するために、いつ、どこで、何をしたらその現象が発生したのかを知ることが重要である。よってそれら背景が何も記載されていない場合には[報告]として十分な情報を有していないと判断し、[報告]属性は付与しない。また、スマートフォンやタブレット端末などの機種名は背景情報として有効であると判断し、機種名が含まれていれば他の記述が無くとも[報告]属性を付与することとした。

例文 1 では、背景は、「アップデした途端」であり、事象は、「スタート画面になった」とこと、「タップしても次に進まなくなった」とこの 2 点であると読み取れるので[報告]タグを付与する。

- 例文 1: アップデした途端スタート画面になってタップして次に進まなくなりました

(2) 要望

利用目的は、主に製品・サービスを提供する側の企画・設計者等が次の商品を作成するに当たり、現在のユーザの要望を参考にすることである。主な[要望]としては既存の商品に対する改善、新しいアイデアの提案などがある。

例文 2 は、新しい機能に対する要望を述べているので[要望]属性を付与する。

- 例文 2: メール機能とかつくて欲しいです

(3) 評価

利用目的は、新規ユーザがレビューされている商品の購入・導入を決定する際の参考情報とすることである。主な[評価]として商品に対する評価や感想がある。

「やらない方がいい」「推奨します」など、新規ユーザに対する助言のような内容、「～できます」といった可能表現、また、他の製品やサービスと比較しているような表現も[評価]に含める。肯定的、否定的の両方を含む。例文 3 では、「楽しく」や「面白いです」といった文面から対象についての感想を述べていると判断できるので[評価]属性を付与する。

- 例文 3: とても楽しく遊んでいます、マルチ面白いです

(4) クレーム

文句、誹謗中傷、不満をまとめて[クレーム]属性とした。例文 4 では、「誤作動が多すぎます」と事実を言っているので[報告]のようにも取れるが、[報告]に必要な背景記述が抜けている為、誤作動が多いことに対する不満である、と解釈し[クレーム]属性を付与する。

- 例文 4: 誤作動が多すぎます

4. 提案手法

レビュー文では、製品・サービスではなく「人」に対して意見が述べられていることがある。人に対する意見とは、例えば他のレビュー投稿者(たち)に対して意見を記述している場合などである。これは、レビュー文が掲載されているサイトが、ユーザ同士の意見のやりとりをするのに利用されることがあるためである。下記に実際の例を挙げる。

- 例文 5: 文句ばっか言う奴はやめちまえ

この文で意見を述べている対象は、「文句ばっか言う奴」であり、商品ではない。このような現象は特に[クレーム]に分類される文に多かった。このようなタイプの文と、評価の対象が商品である通常のレビュー文とを同様に扱くと、商品に対する評価としては適切でない特徴が学習され、意見属性を正しく判定できない可能性がある。そこで、「お前」や「奴」といった人称代名詞が文中で出現したかどうかを示す素性を設定した。

人称代名詞として用いるのは IPAdic¹の代名詞辞書の中に含まれる二人称と三人称である(表 1)。「私」や「僕」などの一人称については、商品に対する意見を自分が述べたものとして強調している場合に使用されることが多かったため、ここで用いる人称代名詞として採用しなかった。

表 1. 人称代名詞として使用した単語

君たち, あいつ, 彼, 貴方, お前, 皆, 皆, 皆んな, そち, 彼女ら, 奴ら, 彼ら, 彼奴, みんな, おまえ, ヤツ, 君, 彼女, やつ, キミ, 奴, てめえ, みなさま, あんた, おめえー, 奴等, みなさん, あなた, 彼等, てめえ, そちら
--

5. 評価実験

5.1 コーパス

評価実験の為にスマートフォン用アプリ配信サイトの売上トップ 10 のゲームアプリのレビューを収集した。2014 年 12 月 25 日と 2015 年 1 月 28 日時点での最新投稿から、各期間 4000 件ずつの計 8000 件のレビューを取得し、3 章で述べた基準で属性を各文に付与した。属性の付与は 2 人のアノテータが実施し、アノテーションの Kappa 係数での一致率は図 1 のようになった。一致率は最も低いもので[クレーム]の約 0.54 であり、他の属性ではすべて 0.6 を超えた。各文のゴールドラベルは 2 人の論理和を設定した。これは今回アノテーションを実施したデータでは、評価以外のラベルの数が少なかったため、ラベルのノイズが上昇しても学習データをより多く確保するために行った。

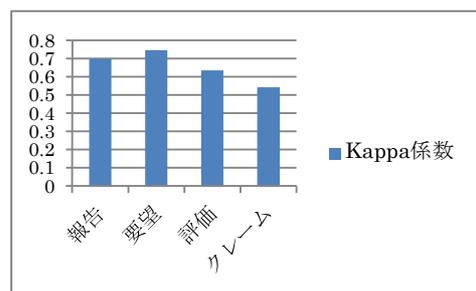


図 1. アノテーション一致率

5.2 実験方法

学習・評価データには 5.1 節で述べた 8000 件のレビューを用い、属性毎にバイナリの判別器を作成した。表 2 に属性ごとの正例と負例の数を示す。文からの単語の抽出には MeCab²を使用した。機械学習手法には Support Vector Machine を用い、その実装には LIBSVM³ をカーネル関数に線形カーネルを設定して用いた。判別性能の評価は 5 分割交差検定で行い、データの分割はレビュー単位で行った。これは、文単位で分割すると 1 つのレビュー内の文が学習と評価に分かれてしまい、不自然な設定になるためである。

判別器で用いる素性には、次の 2 つの組み合わせを用いた。i. Bag of Words 素性(BoW)のみ、ii. BoW と 4 章で述べた人称代名詞の素性(BoW+ND)。また、各属性の性能の下限として属性が必ずあると判別する all-positive ベースライン(Baseline)を計算した。

¹ <https://code.google.com/p/mecab/downloads/detail?name=mecab-ipadic-2.7.0-20070801.tar.gz>

² <https://code.google.com/p/mecab/>

³ <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>

表 2. 学習と評価に使用した正例/負例数

属性	正例数	負例数
報告	735	12570
要望	1574	11731
評価	9210	4095
クレーム	3843	9462

表 3. 実験結果

判定器	人称代名詞	Precision	Recall	F
報告	無し	72.68	26.99	39.10
	有り	72.57	26.85	38.95
	Baseline	5.52	100	10.47
要望	無し	80.24	66.77	72.88
	有り	80.24	66.77	72.88
	Baseline	11.83	100	21.16
評価	無し	79.84	91.88	85.44
	有り	79.87	91.92	85.47
	Baseline	69.22	100	81.81
クレーム	無し	68.87	37.72	48.71
	有り	69.08	38.01	49.01
	Baseline	30.21	100	44.82

6. 実験結果と考察

4つの判定器の結果を表 3 に示す. BoW+ND で[報告]38.95, [要望]72.88, [評価]85.47, [クレーム]49.01 ポイントの精度であった. なお, Baseline の性能は正例と負例の数の比率に依存するため, 属性毎に異なる精度が得られている.

開発者にとって有用なのは特に[報告]に分類される文である. [報告]の評価実験では, Baseline と比べ高い Precision を得ることができたが, Recall が低かった. その要因の 1 つは正例数の少なさと考える. 今後 Recall を改善するためには, 更に多くのレビュー文を収集し, データ量を増やす必要がある.

他のユーザに対する意見を正しく学習するために, BoW+ND で実験をおこなったが, 今回の実験では素性として人称代名詞が出現したかどうかという情報を入れてもその効果が確認できなかった. これは人称代名詞が存在しなくても他のユーザに対する意見を述べている文があることが考えられる. 下記に具体的な例を挙げる.

- 1) 他のユーザを示す表現が明記されていない例
下記の例はいずれも他のユーザに対する不満を述べているものの, ユーザを示す表現が省略されている. したがって人称代名詞の有無では判別できない.
 - ギャーギャー言ってる暇があったら石集めるよ(・▽・)
 - 運営はこんなん書いたところで見ないから文句言ったところで意味ないし w
 - そんなんで評価下げんなや ww
- 2) 他のユーザを指す名詞が人称代名詞ではなかった例
下記の文中で他のユーザを指す語は「方」であるが, これは人称代名詞ではなく形式名詞 (IPAdic では名詞-非自立) である.
 - そういう意見の方他のゲームしてください_(._)_

また, アノテーションで[クレーム]属性が付与された文の中には, 下記の例のように否定的な内容の[評価]に含まれるべき文も混在していた. 図 1 の一致率の調査でも[クレーム]属性が一

番低い値となっている. これがより精度を下げる一因になった可能性がある. 今後はこのような揺れを防ぐために否定的な[評価]と[クレーム]を区別するためのより明確な定義が必要になると思われる.

- 面白いけど、バグが多いです。

7. おわりに

本論文では, 機械学習を用いて利用目的ごとに意見分類する手法を提案した. この手法では, 4つの属性を定義してコーパスを作成し, 人称代名詞が文中で出現したかどうかを示す素性を設定し学習した. しかし, この素性の効果は評価実験では確認することができなかった. 定義した 4属性では, スマートフォン用アプリ配信サイトのレビュー文を対象に, BoW+ND で[報告]38.95, [要望]72.88, [評価]85.47, [クレーム]49.01 ポイントの分類精度を得た.

今回は, 評価文でないものが比較的多く書かれているスマートフォン用アプリ配信サイトのレビュー文を使用し, 利用目的ごとにどの程度分類可能かを調査した. ただし, レビューの対象によって分類基準の細部 (例えば機種名を報告に含めるといった基準) が変更される可能性があるため, 別の対象のレビュー文でも検証し, 有効性を調査するとともに, 共通化を検討する必要がある.

今回作成したコーパスには[報告]に分類されるレビュー文が少なかったため, 他の意見属性に比べて分類精度が低かった. しかし, 学習データをさらに収集することが可能になれば, 他の意見属性と同程度の精度まで向上が見込められる. 今後は学習データを増やすことと, 機種情報などの固有名詞を学習素性に取り入れることによって精度向上に取り組みたい. また, 本研究の目的は開発者が有用な情報を含むレビューを発見し易くすることであるので, 判定器の結果からレビュー文単位で重み付けをし, 優先順位を決める仕組みも検討していきたい.

参考文献

- [Turney 2002] Peter D. Turney: Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews, Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), 2002.
- [Dave 2003] Kushal Dave, Steve Lawrence: Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews, Proceedings of the 12th International Conference on World Wide Web, 2003.
- [飯田 2005] 飯田龍, 小林のぞみ, 乾健太郎, 松本裕治, 立石健二, 福島俊一: 意見抽出を目的とした機械学習による属性-評価値対同定, 情報処理学会研究報告, NL-165, 2005.
- [Stoyanov 2008] Veselin Stoyanov, Claire Cardie: Topic Identification for Fine-Grained Opinion Analysis, Proceedings of the 22nd International Conference on Computational Linguistics (COLING), 2008.
- [藤村 2004] 藤村滋, 豊田正史, 喜連川優: 電子掲示板からの評価表現および評判情報の抽出, 第 18 回人工知能学会全国大会, 2004.