

文章構成を考慮したレビュー P/N 分類手法の提案

The proposal of the P/N classification technique of the review in consideration of text composition.

江川 雄太[†]今野 将[‡]

1. はじめに

インターネットの普及に伴いインターネットで商品を購入する電子商取引を利用する機会が増えている。そのために販売側は商品をインターネット上で販売するためのウェブサイトである EC(Electronic commerce)サイトを用意し、消費者はその EC サイトを利用して商品を購入することが一般的になってきた。日本の電子商取引の市場規模は拡大してきており現在では約 6 兆円もの規模になってきている[1]。

この EC サイトにはただ商品の売買を行うだけでなく、電子掲示板が用意されていてユーザが各商品に対して評価やユーザレビューを投稿することによって、消費者同士の情報交換の場として利用出来るものがある。これは消費者にとってはカタログ情報を見るだけでは得られない商品の情報を得ることのできる場であり、販売側にとっては消費者の生の声を聞く手段のひとつであり、現在では消費者が買い物をする際の情報源として商品进行评估するサイトが他の情報源よりも有効と認識されてきているとの報告がある[2]。

ある商品の購入を検討しているユーザが商品に肯定的なユーザレビューによって購入を決めたり、逆に商品に否定的なユーザレビューによって購入を止めるかもしれない。このように消費者がそのユーザレビューを読むことによってその購買行動に影響を受けるということが考えられるが、具体的にどのような要素がどの程度の影響を与えるかというのは未だ明らかになっていない。

そこで本研究では、ユーザレビューが他の利用者の購買行動に与える影響を明らかにすることを最終目標としたい。しかしながら、ユーザレビューの与える影響を明らかにするためには、ユーザレビューそのものがどのような性質のものであるのか判断できていることが望ましい。具体的には各ユーザレビューが商品に対して肯定的レビュー・否定的レビューのどちらあらかであるのかを判断する手法が必要である。

そこで本研究ではまずユーザレビューを肯定的か否定的かどうかを自動的に分類する手法（以下 PN 分類と呼ぶ）の確立を目的とする。この際に、最終的な目標であるユーザレビューの購買行動に対する影響の分析への利用を考えた場合、ただ単に PN 分類されているだけではなく、そのユーザレビューどのようにその商品进行评估しているのかが判断できる手法であるほうが好ましい。具体的には、同じく肯定的レビューと分類されたレビューでも、全面的に肯定しているレビューと全体的に褒めているが部分的には否定的評価が混ざっているレビューなどがあり、これらを別々に分類したい。

よって本研究ではそのレビューが全体として見て単に肯定的なのか否定的なのかという分類ではなく、レビューを構成する文章を考慮しより細かいクラスまで分類できる PN 分類手法を提案する。

2. 関連研究

PN 分類に関連する先行研究としては、まず藤村ら[3]の研究がある。この研究では肯定・否定的文章に含まれる単語をそれぞれスコアリングし評価辞書を作成し、それを基に文章を肯定と否定に分類している。この単語をスコアリングする手法は、肯定的（否定的）な評判には肯定的（否定的）な概念を持った単語が多く含まれるという仮定から、その単語が肯定的文章に出現する確率と否定的文章に出現する確率の差をとっている。

一方、箴島ら[4]の研究では、藤村ら同様単語の出現確率を利用しているが、分類する際の素性を単語だけでなく単語の組み合わせも利用することで単語だけでは分類が難しい文章であっても分類を可能にしている。また、出現頻度が少ない単語の組み合わせであっても分類に有用な素性が多数存在するという観点から利用している。

また、岡之原ら[5]の研究では、レビューを単に肯定的・否定的のに二値分類するのではなく、どの程度評価しているかを指標として表す手法を提案している。この positive, negative の二極を持った指標を sp-score と定義し、単語の出現回数などから素性の重みを決定し、それを基に SVR (Support Vector Regression) を用いてレビューに対する sp-score を判定している。

これらの先行研究では PN 分類手法としては十分な精度を示しているが、レビュー文内の文章構成を考慮しておらず、同じ点数を付けられたレビューであっても文章の構成によってはレビューの内容に違いがある。本研究ではこれらの先行研究を参考にしつつ、文章構成を考慮した PN 分類手法を提案する。

[†] 千葉工業大学大学院工学研究科電気電子情報工学専攻
Electrical, Electronic and Computer Engineering Speciality,
Engineering Graduate Course Chiba Institute of Technology
Graduate School

[‡] 千葉工業大学工学部電気電子情報工学科
Department of Electrical, Electronic and Computer Engineering
Faculty of Engineering, Chiba Institute of Technology

3. 提案手法

本研究では、ユーザレビューを購買行動への影響分析に利用しやすい様に、肯定・否定の二値分類ではなく、文章構成を考慮したより細かいクラスでの PN 分類を目指している。まず、分類すべきクラスは以下の通りである。

3.1 分類すべきクラス

- (1) 全面的に肯定しているレビュー
- (2) 肯定的なレビューだが、部分的に否定しているレビュー
- (3) 全面的に否定しているレビュー
- (4) 否定的なレビューだが、部分的に肯定しているレビュー

3.2 提案手法

本章では、ユーザレビューを上記の 4 クラスに分類するために、ユーザレビューを文ごとに分類する手法を提案する。これは、肯定的なレビューに混在している否定的表現部分、または否定的なレビューに混在している肯定的表現部分を正しく判断するには文ごとに PN 分類したほうが良いという仮定に基づく。

また、ひとつの文であっても、肯定・否定の両方の表現が混在している場合がある。この場合、肯定的（否定的）表現部分とその逆の表現部分は手がかり語で繋がっていると考えられるので、手がかり語で文を分割しそれぞれ独立した文として扱う。

また、ユーザレビューを構成する各文の分類結果からユーザレビューとしての分類を行う。よって本手法のユーザレビューの PN 分類の流れは以下のようになる。

1. レビューを文ごとに分類
2. 文を手がかり語によって分割
3. 文ごとの PN 分類
4. ユーザレビューの PN 分類

3.3 評価辞書の作成

肯定的または否定的レビューの学習を行う際に、文の PN 分類を行うために用いる評価辞書を作成する。これは各ユーザレビューに出現する単語にスコアリングを行う作業である。このとき肯定的レビューに多く出現する単語は正の値を、否定的なレビューに多く出現する単語は負の値をスコアリングする。

3.4 文の PN 分類方法

学習フェーズで作成した評価辞書を基に、ユーザレビューを構成する文ごとに PN 分類する。この際に、肯定的な文には肯定的表現の単語が、否定的表現には否定的表現の単語が多く出現するであろうとの推測から、評価辞書から取得した文中に出現する単語の値の総和を求め

その総和が正であれば肯定的文章、負であれば否定的文章と判定する。

3.5 ユーザレビューの分類方法

ユーザレビューを構成する各文の PN 分類結果から最終的なユーザレビュー自体の PN 判定を行う。この際に肯定的な文と否定的な文の数の差を取ることでよりユーザレビューを分類する。また、ユーザレビューにおいて文ごとにその重みが違うと考えられる。分類の際の工夫としては、文ごとに重みを持たせることを考えている。EC サイトにはユーザレビューごとにレビュータイトルが付属しているものが多い。このレビュータイトルはそのレビュー文章を短く要約したものである場合が一般的であるから、肯定的な文・否定的な文の両方を含むレビューでもレビュータイトルと一致する内容を述べている文があった場合に大きな重みを持たせてやれば良い。こうして重み付けをした肯定的・否定的文の差をとりその値の正負で最終的なユーザレビューの分類を行う。

4. まとめ

本研究では、レビュー中の文章構成を考慮した PN 分類手法の提案を行った。今後は提案手法をもとに PN 分類システムを実装し、分類実験を行い、提案手法の有用性を確認したい。この際に、単語のスコアリングの方法や文の重み付けなどの分類手法の見直しを図りたい。また、最終的には提案する PN 分類手法を用いてユーザレビューの消費者の購買行動への影響分析への利用へと繋げていきたい。

参考文献

- [1] 経済産業省：平成 20 年度我が国の IT 利活用に関する調査研究，
<http://www.meti.go.jp/press/20091014003/20091014003.html>
- [2] 佐々木裕一：商品購買における評価サイトの有効度，情報メディア研究 3(1)[2005]
- [3] 藤村滋，豊田正史，喜連川優，Web からの評判および評価表現抽出に関する一考察，情報処理学会研究報告，Vol.2004，No.72，pp. 461--468. 電子情報通信学会技術研究報告，Vol.104，No.176，pp. 141-146.
- [4] 箴島侑子，嶋田和孝，遠藤勉，系列パターンを利用した評価表現の分類，言語処理学会第 11 回年次大会，2005.3
- [5] 岡野原大輔，辻井潤一，評価文に対する二極指標の自動付与，言語処理学会第 11 回年次大会発表論文集，pp. 664-667，March 2005