Vol.2018-MPS-121 No.12
2018/12/18

有用な商品レビュー抽出のための過賞賛レビュー除去フィルターの性能評価

草刈祐子[†1]　鈴木孝彦[†2]　廣川佐千男[†2]

概要：オンラインショップで買い物をする際，商品につけられているレビューは購入判断の重要な要素になる．しかし，これらのレビューの中には手放しで商品を褒める客観性のないレビューや，やらせのレビューなどの役に立たないレビューも多く存在する．それらのレビューは購入判断の妨げになる．我々は，それらの役に立たないレビューを除外するフィルターを作成し，性能を評価した．フィルターの作成にあたり，商品に対するレビューの評価分布から，過賞賛レビューを抽出した．最高評価レビューが一定以上の割合で投稿されている商品に着目した．

キーワード：テキストマイニング，機械学習

Performance evaluation of over-praise reviews removal filter for useful product reviews extraction

YUKO KUSAKARI[†1]　TAKAHIKO SUZUKI[†2]
SACHIO HIROKAWA[†3]

Abstract: When shopping at the online shop, reviews attached to products are important elements of purchase judgment. However, some of these include reviews that do not objectiveness that praise the item, such as fake reviews etc. Many useless reviews also exist, and these reviews hinder purchase decisions. We created filters to exclude those useless reviews and evaluated the performance. When creating the filter, evaluate the review of the product Extracted praise reviews were extracted from the distribution, focusing on products that have been posted at a certain rate or higher.

Keywords: Text mining, Machine learning

1. はじめに

オンラインショップで商品を購入する場合，商品のレビューは購入の決定に役立つ．しかし，これらのレビューの多くには，有益な情報なしで製品を賞賛するものが含まれる．このようなレビューを迷惑メールフィルタのような仕組みで除外できれば，オンラインショップの利用者にとっては有益である．

あるレビューが「褒めすぎレビュー」であるか否かを教師付き機械学習で判別するためには，褒めすぎレビューの正例が大量に必要である.しかしながら，トレーニングデータを準備することは必ずしも容易ではない．

本稿では，商品に対するレビューの評価分布から，褒めすぎレビューのトレーニングデータを抽出し，それらのレビューを取り除くことができるフィルターの構築と評価を行った．その際，楽天市場の商品レビューデータについて，最高評価である星 5 レビューの割合が非常に多く投稿されている商品に着目した．そのような商品に対する星 5 レビューを褒めすぎレビューとみなし，SVM を用いた機械学習による判別および特徴語の抽出を試みた．

2. 関連研究

Kokkodis ら[1]は，Amazon レビューのいくつかのジャンルについて，確認済購入者(Verified Purchaser)によるレビューと，それ以外のレビューを比較して，レビューの特徴と有用度を分析した．林ら[2]は楽天レビューの「ダイエット・健康」「おもちゃ・ホビー・ゲーム」「パソコン・周辺機器」の 3 つのジャンルについて，未購入者のレビューに注目し特徴と有用性を分析した．草刈ら[3]は，楽天レビューの「ダイエット・サプリ」ジャンルについて，5 点評価のレビューの割合が高い商品を収集し，相当する 5 点レビューの特徴を目視で分析した．ジャンルを限定せずに，5 点レビューが 7 割以上の商品に対するレビューを褒めすぎレビューととらえ，その特徴を分析した草刈ら[4]は，[3]の条件を満たすレビューを正例とし，SVM+属性選択[5]による識別実験を行なった．結果，一定以上の判別性能を示したが，抽出された属性の中に，商品ジャンルに関係するものが多く見られ，レビューされた商品のジャンルの属性が強く影響している可能性があった．

本研究では，商品のジャンル属性がどの程度「褒めすぎ

†1 九州大学 システム情報科学府
Faculty of Information Science and Electrical Engineering, Kyushu University
†2 九州大学 情報基盤研究開発センター
Research Institute for Information Technology, Kyushu University

ⓒ 2018 Information Processing Society of Japan

レビュー」の判別に影響するかを明らかにする．本稿は以下，次の構成をとる．3 節では実験に利用した商品レビューデータを説明する．4 節では本研究における「褒められすぎ商品」と「褒めすぎレビュー」の定義を示す．5 節では草刈ら[4]の結果を検証し，ジャンル属性が判別性能にどの程度影響しているのかを調べる．6, 7 節ではジャンルの影響を取り除いた実験とその結果．8 節は今後の課題である．

3. レビューデータ

本研究では，楽天株式会社が国立情報研究所の「情報学研究レポジトリ」[6]を通じて研究者に提供している楽天データセット中の，楽天市場商品レビューデータ[7]（以後楽天レビューデータ）を対象とする．表 1 に使用したデータの項目を示す．

表 1　今回使用したデータの項目

ラベル	説明
商品 ID	商品毎に固有の英数字文字列
ジャンル ID	ジャンルごとに固有の数字列
評価点数	星 1~5 の 5 段階評価．
レビュー本文	**主な分析対象**
レビュー日時	レビューの投稿された日時
その他の項目	今回は使用しないため省略

楽天データの商品ジャンルは木構造の親子関係になっている．大きな商品ジャンル(「家電」など)内により詳細なジャンル(「キッチン家電」「季節家電(冷暖房)」など)，さらに小さなジャンル(「電子レンジ」「除湿機」など)と入れ子になっている．

4. 褒められ過ぎ商品と褒め過ぎレビュー

1 ヶ月間に 10 件以上のレビューがある商品のうち，星 5 レビューの割合が 7 割を超えている商品(図 1 の商品 A, B)を「褒められすぎ商品」と呼ぶ．褒められすぎ商品の星 5 レビュー(図 1 の実線枠内)を「褒めすぎレビュー」と呼び，正例として使用する．また，SVM で機械学習を行うためには，褒めすぎレビューの負例が必要である．そこで，一月に 10 件以上のレビューがある商品のうち，星 5 の割合が 3 割を下回っているもの(図 1 の商品 E, F)の星 5 レビュー(図 1 の青点線枠内)を負例として使用する．

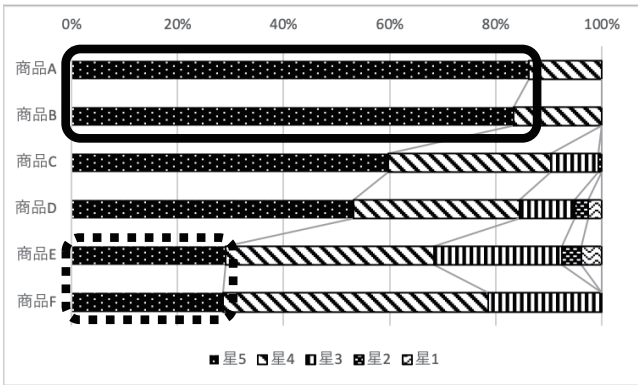

図 1 商品ごとの評価の割合

5. 予備実験

[4]ではレビュー数 100 件以上の商品を対象としているが，6 節における実験で対象となる商品数を確保するため 10 件以上に変更している．1 ヶ月ごとのデータについて，正例，負例レビューを抽出．対象の商品とレビューの数を表 2 に示す．レビュー本文を Bag of Words でベクトル化し，SVM+属性選択を用いて学習した 5 分割交差検定で性能を判定した．

表 2　実験対象の商品とレビュー数

		レビュー数	商品数
1 月	正例	123448	4653
	負例	32085	1258
2 月	正例	105294	4052
	負例	29842	1295
3 月	正例	107969	4212
	負例	31142	1381

次に，ジャンルがどの程度判別に影響しているのかを見積もるため，商品のジャンル ID(および商品の親ジャンル ID)を Bag of Word に追加し，5 分割交差検定を行った．

図 2 のグラフの模様部分は，商品ジャンル ID を特徴語として使用しなかった場合の Precision, Recall, F-measure である．商品ジャンル ID を特徴語に追加したことで，グラフの白い部分は性能が向上，黒い部分は下降した．表 3 には上位の特徴語 100 のうちジャンル ID の数を示す．ジャンル ID を追加したところ，判別性能は向上し，ジャンル ID が特徴語の上位を占めた．これは，商品ジャンル ID が判別に大きく影響することを示す．

ⓒ 2018 Information Processing Society of Japan

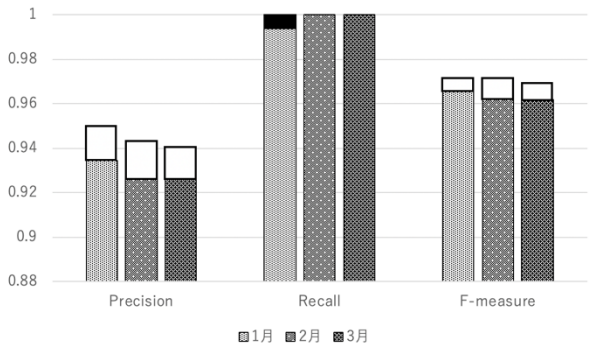

図 2 2010 年 1 月~3 月の全てのジャンルでの識別性能

表 3 各月の特徴語上位に含まれる商品ジャンル ID の数

		上位 10	上位 100
1 月	正例	9	96
	負例	10	100
2 月	正例	10	99
	負例	10	100
3 月	正例	10	100
	負例	10	100

図 2 より一定以上の識別性能を得られたことがわかったが, また表 3 に示す通り正例負例とも特徴語の上位にジャンル ID が多く含まれることもわかった.
このことから, レビュー本文の内容よりも商品ジャンル ID が識別性能の向上に有効であることがわかった.

特定のジャンルに限定して実験をすることでジャンルに関する特徴語を除くことができるのではないか, と考え, ジャンルを絞って実験をした.

6. ジャンルごとの実験

5 節で行った予備実験では, ジャンルを識別することで[4]と同等の結果を得た.

ジャンルに関連する属性を除去するために, 特定のジャンルにおけるレビューに限定して, 褒めすぎレビューの判定を行うこととした.

6.1 実験方法

ジャンルを限定した場合, 褒められすぎ商品数および褒めすぎレビュー数が少なくなり, SVM による学習が困難になる. そのため, 以下の基準に従って, 商品数・レビュー数がある程度多くジャンルの属性の影響が強く出ている部分を対象とした.

・全ジャンルで実験した際に子供ジャンルが特徴語上位 20 位以内に存在しているジャンル
・最上位のジャンル(親が存在しないジャンル)
・子供ジャンルが正例と負例両方の特徴語上位に存在しないジャンル

上記の条件に適合するジャンルは 5 ジャンル存在し, そのうち商品数が上位 2 番目までのジャンル G1:2010 年 1 月「美容・コスメ・香水」および G2:2010 年 1 月「家電」ジャンルについて実験を行なった. 2 つのジャンルに含まれる商品数, レビュー数(正例・負例)を表 4 に示す.

他の 3 ジャンルについては正例の商品数・レビュー数が少ないため, 実験を行わなかった.

また, 正例と負例のレビューの条件は 4 で定めたものを使用した. 以下にレビューの内訳を示す.

表 4 対象ジャンルの商品数およびレビュー数

ジャンル ID	分類	商品数	レビュー数
G1	正例	627	15898
	負例	107	1663
G2	正例	55	960
	負例	79	1974

G1:「美容・コスメ・香水」ジャンル
G2:「家電」ジャンル　　どちらも 2010 年 1 月分

6.2 識別性能

G1, G2 における, 特徴語にジャンル ID を含まない場合の識別性能を表 5 に示す. 参考として[4] および 5 節の予備実験の識別性能を併記する.

表 5

	Precision	Recall	F-measure
G1	0.8796	0.9847	0.9291
G2	0.7883	0.8741	0.8285
予備実験	0.9346	1.0000	0.9662
[4]の実験	0.9784	0.9618	0.9700

G1,G2 とも Recall が高く, Precision が低い. 5 節の予備実験では正例・負例の数が他の実験より 1 桁以上多く, 各値の直接比較は難しい.

6.3 ジャンル ID の識別性能への影響

ジャンル ID（対象ジャンル中の子ジャンルの ID）を属性に追加して識別実験を行ったところ, 図 3, 4 に示す結果を得た. ジャンル ID を追加による識別性能の変化にパターンは見られない.

ⓒ 2018 Information Processing Society of Japan

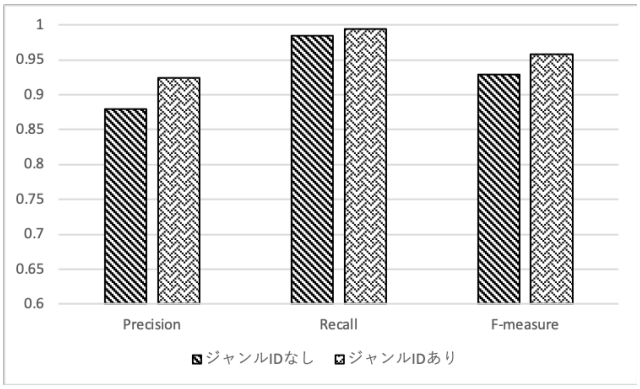

図 3 属性選択数 100 の場合のジャンル ID の有無による性能比較(2010 年 1 月のジャンル G1)

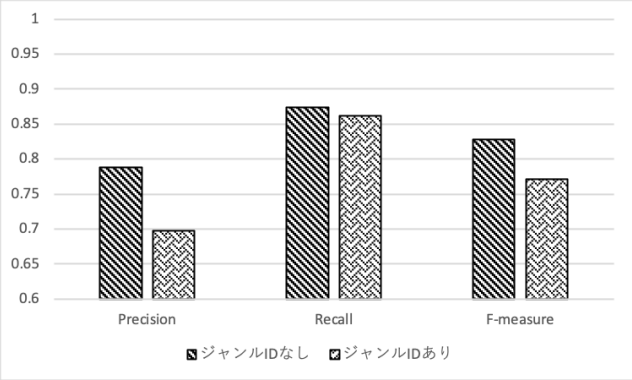

図 4 属性選択数 100 の場合のジャンル ID の有無による性能比較(2010 年 1 月のジャンル G2)

6.4 特徴語

ジャンル ID を属性に含まない場合, SVM+属性選択の正負特徴語各上位 10 を以下の表 6 に示す.

表 6 2010 年 1 月のジャンル G1 の特徴語
(ジャンル ID を特徴語に含まない)

	正例	負例
1	みなさん	コロコロ
2	清潔	カバー
3	あがり	染め上がり
4	経済	グレープ フルーツ
5	油	かかる
6	行っ	テレビ
7	袋	染まっ
8	保湿	素手
9	多少	優しい
10	立ち	あまりに

表 7 2010 年 1 月のジャンル G2 の特徴語
(ジャンル ID を特徴語に含まない)

	正例	負例
1	安	いろんな
2	多い	かわいく
3	又	キレイ
4	精米	性能
5	なかなか	無料
6	として	サイクロン
7	ストレート	安心
8	ˆ	持ち運び
9	お願い	ペットボトル
10	癒さ	まず

6.5 ジャンルの影響

ジャンル ID を属性に加えた場合, 正負特徴語の中に含まれるジャンル ID の数を表 8 に示す. 多くのジャンル ID が特徴語に含まれるが, 予備実験での表 3 より比率は少なくなっている.

表 8 特徴語上位に含まれるジャンル ID の数

		上位 10	上位 100
G1	正例	8	50
	負例	6	43
G2	正例	7	57
	負例	10	43

7. 考察

特定のジャンルのレビューについて「褒め過ぎレビュー」の識別をした結果, 一定の識別性能が得られた. ジャンルを限定しない場合よりも識別性能が下がっているが, この理由は,

・学習例が減少したため
・ジャンルを属性として識別することが困難になったため

の可能性がある.

特定のジャンルについての識別では, Recall は比較的高かったが, 先行研究[4] に比べて Precision の値が低かった. つまり, 「適切な高評価レビュー」が「褒めすぎレビュー」と識別される場合がある程度存在する.

この原因について, 以下のことが指摘できる. 3 の定義では, 「褒められすぎ商品」のすべての星 5 レビューを「褒めすぎレビュー」としているが, 該当商品の星 5 レビューの中には, 適切な高評価レビューが含まれている場合があ

ⓒ 2018 Information Processing Society of Japan

り，そのような適切な高評価レビューが「褒めすぎレビュー」の正例として含まれる．

全てのジャンルを対象とした 5 節の実験では，ジャンル ID を属性に追加することによって Precision が向上したが，特定のジャンルを対象とした 6 節の実験では，必ずしもそのような傾向は見られない．

実験を行った２つのジャンルでは，SVM+属性選択の特徴語が全く異なっている．これは，「褒めすぎレビュー」の特徴はジャンルごとに異なり，識別のためには各ジャンルに応じた学習が必要なことを示唆する．

ジャンル ID（対象ジャンル中の子ジャンル ID）をレビュー属性に追加した実験で，上位 100 の特徴語にジャンル ID が 43~57%含まれた．しかしながら，ジャンル ID を追加するとかえって識別性能が下がる場合があり，ジャンル ID が識別に有用に働くとはいえない．

また，ジャンル ID を属性に含まない実験では，ジャンルを子ジャンルに分割するような特徴語は上位特徴語のごく一部である．このことから，対象を特定のジャンルに限定した，褒めすぎレビューの学習は有効であると言える．

8. 今後の課題

本稿では，限定されたジャンルにおける一部のレビューについての実験を行った．一般のジャンルで実験を行うために，そのジャンルにおける「褒めすぎレビュー」の正例を抽出する必要がある．ジャンル内に 4 節の条件に適合するレビューが少ない場合の対処法は今後の課題である．

7 節で指摘したように，「褒められすぎ商品」の星 5 レビューを全て褒めすぎレビューとしているため，実際には褒めすぎレビューではないものも正例に含まれていると予想される．これについては，他の条件と組み合わせることで，適切な正例を求めることを考えたい．

本稿では，限定されたジャンルにおける一部のレビューについての実験を行った．一般のジャンルで実験を行うためには，そのジャンルにおける「褒めすぎレビュー」の正例を抽出する必要がある．

また，本研究の方式で学習した SVM が，4 節の条件を満たさない一般のレビューについて「褒めすぎレビュー」の識別に有用か否かを確認する必要がある．

さらに，単純な星 5 割合以外にも統計的手法[8]やセンチメント分析[9]，を利用する手法も可能であるため，そちらも検討する．

参考文献

[1]Kokkodis, M., Lappas, T., The relationship between disclosing purchase information and reputation systems in electronic markets, ICIS 2016

[2]林驍， 伊東栄典， 廣川佐千男， 大規模 EC サイトの商品レビュー傾向分析， 情報処理学会研究報告， Vol. 2015-ICS-181, No. 7， 2015

[3]草刈祐子，林驍，鈴木 孝彦，伊東 栄典，廣川 佐千男，有用な商品レビューの特徴と抽出について，情報処理学会研究報告，火の国シンポジウム, B3-4, 2018

[4]草刈祐子，鈴木孝彦，廣川佐千男，中藤哲也 過賞賛のレビューの検出と特徴語抽出，第 17 回情報科学技術フォーラム, F-016, 2018

[5]Toshihiko Sakai, Sachio Hirokawa, Feature Words that Classify Problem Sentence in Scientific, IIWAS'2012 Special Issue Journal: Intl J of Data Warehousing and Mining (IJDWM), 2012

[6]NII， 情報学研究リポジトリ, http://www.nii.ac.jp/dsc/idr/ (accessed at Jan.19.2018).

[7]NII， 楽天株式会社， 楽天データセット, http://www.nii.ac.jp/dsc/idr/rakuten/rakuten.html. (accessed at Jan.19.2018).

[8]中藤哲也，草刈祐子，鈴木孝彦，廣川佐千男，統計的性質を用いた通販サイトにおけるレビュアの評価の試み，電子情報通信学会技術研究報告，人工知能と知識処理研究会(AI), ISSN 0913-5685, 2018

[9] Lionel Martin, Pearl Pu, Prediction of helpful reviews using emotions extraction, AAAI'14 Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, Pages 1551-1557

ⓒ 2018 Information Processing Society of Japan