

Amazon レビューを用いた有用性の判定実験

The judgment experiment of the usefulness among Amazon's review

倉橋 宏幸† 青野 雅樹†
 Hiroyuki Kurahashi Masaki Aono

† 豊橋技術科学大学 大学院 情報・知能工学専攻
 Dept. of Computer Science and Engineering, Toyohashi University of Technology

1. はじめに

SNS やブログなどの利用者の書き込みには、商品や映画について、その感想などを記述した評判情報が多く含まれている。これら Web 上に大量に蓄積されている評判情報は、出版物などと比べると品質の程度が低く、有用な部分を判断した上で利用することが理想的である。本研究では、評判情報が多数含まれているレビューサイトに着目する。

レビューサイトに投稿されている評判文書には、ユーザーにより「参考になったかどうか」の評価値が付与されている。これを利用し、多くのユーザーから支持を得ている評判文書を有用性が高い文書であるとした上で、その特性の分析と共に有用性の判定を行う手法を提案する。

2. 関連研究

レビューの有用性に着目した研究は過去にも行われている。山澤ら[2]の研究は、ユーザーが内容を信用して利用できるレビュー文(有用文)を自動抽出することを目的としている。有用文であるかは人手でタグを付与し、形態素情報のみを用いた SVM による有用文分類実験を行っている。小倉ら[3]の研究では、レビュー情報に評価が付与されていることに着目し、レビュー品質の考察を行っている。それに対し本研究では、Amazon[1]レビューに着目し、有用性の判定実験を行う。具体的には、レビューに付与されている「参考になったかどうか」の指標から有用タグを付与し、機械学習を用いた有用性の判定実験を行うものである。今回の実験では、Amazon レビューのみを用いて判定を行うが、レビュー形式に依存しない有用性の判定を目指し、特性の分析も行っていく。

3. Amazon レビュー文の有用性判別

本研究ではレビューの有用性を考慮するため、梅村ら[4]の手法を引用し、支持率を以下のように定義する。

Amazon などのレビューには「参考になったかどうか」の評価値が付与されている。あるレビューに対して行われた全ての投票数を N 、レビューに対して同意するに投票された数を Y としたときの支持率 f を式 (1) にて定義する。

$$f = \frac{Y}{N} \quad (1)$$

式 (1) に示す値が高ければ、多くの利用者から同意を得ていることを示し、レビューの有用性が高く、高品質なレビューであるということを表している。

3. 素性設計とスクリーニングについて

有用性の判定には、提案素性 (全 25 素性) と、形態素による BOW (Bag Of Word) を用いる。まず、表 1 に提案素性の内容を示す。ここで極性スコアとは、高村らの単語感情極性対応表[5]を用いており、文書が positive/negative どちらの傾向を持つかを数値化したものである。

表 1. 提案素性

- | |
|--------------------------------------|
| 1: レビュー文内の文字数 |
| 2: レビュー文内の名詞の出現回数 |
| 3: レビュー文内の形容詞の出現回数 |
| 4: 最初に投稿されたレビューとの日付の差 |
| 5: 投稿順位 (日付に基づくランキング) |
| 6: レビューがアイテムに付与した点数 |
| 7: アイテムの平均点とレビューがアイテムに付与した点数との差 |
| 8: レビューの極性スコア |
| 9: アイテムの極性スコアの平均とレビューの極性スコアとの差 |
| 10,11: レビュータイトルの形容詞の出現回数とその有無 |
| 12,13: レビュータイトルの名詞の出現回数とその有無 |
| 14,15: レビュー文内のリンクの出現回数とその有無 |
| 16,17: レビュー文内の段落の出現回数とその有無 |
| 18,19: レビュー文内の改行の出現回数とその有無 |
| 20,21: レビュー文内の助動詞「です」、「ます」の出現回数とその有無 |
| 22,23: レビュー文内の感嘆符「!」の出現回数とその有無 |
| 24,25: レビュー文内の疑問符「?」の出現回数とその有無 |

また、BOW に含まれる形態素について、素性のスクリーニングを行った。具体的には、内容語 (名詞・動詞・形容詞・副詞) を対象として、高品質/低品質なレビューそれぞれについての出現頻度の偏りを基にしてスクリーニングを行い、BOW (全 5392 素性) を作成した。

4. 実験

4. 1 データセットの概要

本実験では、Amazon ベストセラーの中から、DVD カテゴリより 2001 年から 2011 年の 11 年間について収集した

データを用いる。レビュー総数 70991 件より、実験のために用いるデータの選定を行った。その手順を以下に示す。

- ① レビューを参照した人数を考慮するため「参考になったかどうか」の投票数が 10 件未満のものを除外する。
- ② 式 (1) から高品質/低品質のタグを付与する。

$\left\{ \begin{array}{l} \text{支持率0.7以上のレビュー} \cdots \text{高品質} \\ \text{支持率0.3以下のレビュー} \cdots \text{低品質} \end{array} \right.$

⇒支持率データ集合とする。

- ③ 支持率データ集合より、高品質/低品質なレビューそれぞれ 500 件ずつランダムで抽出する。

以上の手順により選定された 1000 件のデータを実験に用いる。

4. 2 実験概要

高品質/低品質のタグを付与したレビューデータの分類実験を行う。分類器には、SVM と RF (Random Forest) の二つを利用し、提案素性、及び BOW を学習する。分類精度の比較のため、提案素性のみ、BOW のみ、提案素性に BOW を加えたものの 3 パターンで評価を行った。評価は 10 分割交差検定で行うものとし、指標としては Accuracy, Precision, Recall, F-measure を利用する。

また、RF にて素性の重要度を求める。このことで分類に有効な素性が分かり、レビューの支持率に影響する文書の特性を獲得できる。

5. 実験結果

実験によって得られた SVM での評価を表 2, RF での評価を表 3 に示す。評価の結果、SVM 及び RF ともに提案素性のみで分類を行ったものが最も高い評価であった。これは、アイテムによってレビューに出現する形態素が大きく異なっており、BOW を用いた場合では判別が難しいことが原因であると考えられる。今回の実験では、形態素の出現の有無 (boolean) で BOW を作成したが、これに何らかの工夫を加えて分類に用いる必要がある。

表 2. SVM での評価

素性	Accuracy	Precision	Recall	F-measure
提案素性のみ	0.738	0.766	0.725	0.745
BOW のみ	0.634	0.58	0.650	0.613
提案素性+BOW	0.635	0.594	0.647	0.619

表 3. RF での評価

素性	Accuracy	Precision	Recall	F-measure
提案素性のみ	0.753	0.742	0.759	0.750
BOW のみ	0.703	0.696	0.706	0.701
提案素性+BOW	0.750	0.738	0.756	0.747

また、RF にて素性の重要度を求めたところ、上位 20 素性に提案 25 素性中の 13 素性を包含する結果が得られた。結果の上位 10 素性を表 4 に示す。

表 4. 素性の重要度

順位	素性
1	アイテムの平均点とレビューがアイテムに付与した点数との差
2	アイテムの極性スコアの平均とレビューの極性スコアとの差
3	レビューがアイテムに付与した点数
4	レビュー文内の文字数
5	レビュー文内の名詞の出現回数
6	レビューの極性スコア
7	投稿順位 (日付に基づくランキング)
8	レビュー文内の助動詞「です」、「ます」の出現回数
9	最初に投稿されたレビューとの日付の差
10	レビュー文内の改行の出現回数

レビューが positive/negative どちらの傾向を持つかを表す素性が上位にくる結果となった。レビューに対して行われる投票は、レビューが同意できるか否かを判断して行われているものであり、アイテムにより求められる表現が異なるためと考えられる。また、文字数や品詞などのレビュー文内から得られる素性だけでなく、投稿された日付に基づく素性も上位に現れる結果となった。

6. おわりに

本研究では、レビューに付与されている「参考になったかどうか」の指標から有用タグを付与し、機械学習を用いた有用性の判定実験を行った。学習する際の素性を提案し、その有効性の確認も行った。

実験結果では、BOW を用いた場合よりも提案素性のみで分類を行ったものが高い精度を得た。また、素性の重要度では、特にレビューが positive/negative どちらの傾向を持つかを表す素性が上位にくる結果となった。

今後の課題として、有用なレビューの特徴を具体的に整理し、Amazon 以外のレビュー形式にも対応できるようにしたい。また、高品質/低品質の 2 値判定ではなく、更に詳細な分類、ランキング学習などにも発展させていきたい。

参考文献

- [1] Amazon.co.jp, <http://www.amazon.co.jp/>
- [2] 山澤美由起, 吉村宏樹, 増市博 “Amazon レビュー文の有用性判別実験”, 情報処理学会研究報告, 2006
- [3] 小倉達矢, 穴戸開, 今藤紀子, 山口実靖, 浅谷耕一 “レビューサイトにおける良質なレビューの特性とそれを考慮した評判情報の抽出に関する一考察”, 電子情報通信学会第 19 回データ工学ワークショップ論文集(DEWS-2008), 2008
- [4] 梅村和宏, 鈴木優, 川越恭二, “批評文の信頼度を用いたレビュー対象検索手法”, 電子情報通信学会第 19 回データ工学ワークショップ論文集(DEWS-2008), 2008
- [5] 高村大也, 乾孝司, 奥村学, “スピンモデルによる単語の感情極性抽出”, 情報処理学会論文誌ジャーナル, Vol.47 No.02 pp. 627--637, 2006.