

文書分類を基にした Web 上の評判抽出に関する一考察

A Consideration of Extracting Reputations on the Web Based on Document Classification

藤村 滋[♥] 豊田 正史[♦]
喜連川 優[♦]

Shigeru FUJIMURA Masashi TOYODA
Masaru KITSUREGAWA

Web上の膨大な情報の中から評判を自動的に抽出することで、企業においてはマーケティングやクレーム処理、個人においては意思決定支援等への応用が期待されている。しかし、評判は主観的な情報であり、その抽出にはテキストの意味を取り扱う必要性が生じ、容易には抽出ができない。そこで、本報告ではあらかじめ収集した肯定的な評判と否定的な評判をコーパスとし、このコーパスから統計的に評価表現を抽出する手法を示す。次に、肯定・否定への文書分類を基にして、評判の抽出へと拡張する手法について紹介する。また、一連の手法を用いて Web からの評判抽出を行う簡単なシステムを構築したので紹介する。最後に、その結果得られた課題についての考察を報告する。

Automatic extraction of reputations from huge information on the Web is much attended. Extracted reputations can be utilized for marketing, claim management in companies and decision support in customers. But, because of subjective aspects of reputations, extracting reputations is not so easy. This paper describes a reputation classifying method based on statistic extracting evaluative keywords. Then, extracting opinions which include in reputation can be also realized by expansion of this method. We also introduce our system of extracting reputation. In the process of making this system, it turns out that there are various problems. So, we make a report of consideration about these problems.

1. はじめに

近年、ブロードバンドの急速な普及とともに、個人がWeb上で評判を検索する事が一般化している。個人の注目が集まることで、Webコミュニティの社会的影響力は加速度的に増大し、Web上の評判を認識しておくことの重要性は急速に高まっている。今後、Web上の情報が企業において適切なマーケティング戦略を構築する上でさらなる重要性を増してく

[♥] 学生会員 東京大学大学院情報理工学系研究科修士課程
fujimura@tkl.iis.u-tokyo.ac.jp

[♦] 正会員 東京大学生産技術研究所
[fujimura@tkl.iis.u-tokyo.ac.jp](mailto:{toyoda, kitsure}@tkl.iis.u-tokyo.ac.jp)

るのは明白である。[1]

しかし、実際にWebで評判を調べる作業は、膨大な量のテキストを読む必要性を含み、読むべきテキストの量的な側面だけでなく、時間的な側面からも困難な作業となっている。以上のような背景から、個人においては商品購入等の意思決定支援を容易にするため、また企業においてはマーケティングやクレーム処理の支援、および費用の削減のために、評判を自動的に抽出する手法に対する期待が高まっている。また、抽出された評判を上記のような目的で利用することを考慮すると、肯定・否定に分類されていたほうが、その利用価値が高い。

本稿では肯定・否定への文書分類を基にしたWebからの評判抽出に関する実験について報告する。また、実験から得られた課題に対する考察についても報告する。以下、2章では関連研究について述べる。3章では評判の肯定・否定による文書分類に関して本報告で用いた手法について述べ、その評価実験について記す。次に、4章でWebからの評判抽出システムの構築、現状での問題点およびその考察について述べ、最後に5章で本報告のまとめについて記す。

2. 関連研究

評判の抽出に関する先行研究としては、立石[2]らの研究があげられる。この研究では、ユーザが入力した商品名とあらかじめ辞書として用意した評価表現を近接演算する方法を用いて、インターネットの Web ページから意見を抽出している。また、抽出した意見の意見らしさ（適性値）を構文的な特徴を利用して判定している。しかし、この研究では評価表現辞書の作成、適正值判定処理どちらもヒューリスティックに構築されていた。評価表現は話題のドメインによって大幅に変わる。ドメインごとのヒューリスティックな評価表現辞書の作成は容易ではなく、また登録されていない表現は評判として抽出されることがないという問題点がある。

一方、Web上のレビューを肯定・否定に分類し、抽出を行った例としては、Dave[3]らの研究がある。しかし、この研究での対象言語は英語のみであった。そこで、日本語でもこの手法が応用可能か確かめるため、今回の報告ではこの論文の手法を参考にして実験を行った。この研究における手法の詳しい説明については、本報告で用いた手法も含め次章で報告する。

この他に評判に関係する研究としては、立石らの研究での評価表現辞書の強化を試みた小林ら[4]の研究があげられる。また、従来からの機械学習が評判の肯定・否定の分類にどの程度有効であるかを確かめた Pang[5]らの研究や、WSJ(Wall Street Journal)の記事を事実と意見に分類し、かつ意見を肯定・否定に分類することを試みた Yu[6]らの研究もある。さらに、リサイクルに関する新聞記事から、リサイクルに望ましいことを表す表現（望ましくない表現）をブートストラップ的に獲得することを試みた乾ら[7]の研究があげられる。

3. 文書分類を用いた評判抽出

本章では、まず一連の処理の中心となる評価表現辞書の作成法および評判の肯定・否定への文書分類法について述べ、次に文書分類に関する評価実験について報告する。

3.1 評価表現辞書の構築

● 訓練コーパス

今回の報告で実際に取り扱うドメインとして“ノートPC”を選んだ。評価表現辞書を統計的に作成するための訓

練コーパスとしては、価格.comのノートPCに関する掲示板の2003年度の書き込みを用いることにした。肯定的な評判935件、否定的な評判551件である。価格.comでは書き込みを行う人が、使用レポート(良)、使用レポート(悪)というタグをつけることができ、これを肯定的な評判および否定的な評判とみなし、コーパスとして利用した。

● 属性の選択

評価表現の属性選択の手法としては、次の3種類の手法を試した。ひとつは(1)形容詞、形容動詞のみを属性とする手法であり、ふたつめは(2)名詞、未知語のみを属性とするという手法である。そして、最後は(3) (1)+(2)の全ての属性を選択した手法である。形態素解析には、茶筌を用いた。

形容詞・形容動詞については、主に日本語でモノの評価を表す表現であるので属性として採用した。一方、名詞・未知語は従来からのトピック主体の文書分類で主要属性として採用されてきた。評判のPN分類¹のように主観的な情報主体の文書分類で名詞・未知語がどのような影響を与えるのか調べる意味でも属性として採用することとした。また、名詞を属性として採用することにより、「満足(サ変接続名詞)」や「最高(一般名詞)」などの評判に大きな影響を与えると考えられる語を取り込むことができると期待される。

未知語については、例えば、「Pentium4」「IEEE1394b」などのようにPCのスペックに用いられる語が未知語として評価表現に取り込まれ、その語が肯定的か否定的かを見ることでマイニングに繋がれると考えている。

● スコアリング手法

肯定的(否定的)な評判には、肯定的(否定的)な概念を持った語が多く含まれているはずである。この仮定を元に、肯定的な評判と否定的な評判の差をとる。一般的な語はどちらの文書にも同様に出現するはずであるから、その影響は打ち消される。評判において特徴的な語が肯定的な評価表現については正の値をもって、否定的な評価表現については負の値をもって抽出される。

実際には、次のような式でスコアリングを行っている。

$$score(w_i) = \frac{P_p(w_i) - P_n(w_i)}{P_p(w_i) + P_n(w_i) + k} \quad (-1 \leq score(w_i) \leq 1) \quad (1)$$

ここで、 $P_p(w_i)$ は肯定的な評判で属性 w_i が出現する確率である。同様に $P_n(w_i)$ は否定的な評判でのそれである。また k は、例えば $P_n(w_i)$ が0であった際に、 $P_p(w_i)$ が0.1でも0.8でも結果としてスコアが1となってしまうという、1/1の問題を解決するために分母に加えた実数である。

最後に、このスコアリングによって高いスコアを獲得した属性の例を、表1に示す。

3.2 PN分類手法

今回試したPN分類法については、式(2)、(3)に示す。

$$Score(d) = \sum_{ALLw_i} score(w_i) \quad (2)$$

¹ 以下、肯定・否定分類のことをPN分類と記す

表1 スコアの大きな属性

Table 1 Hi-scored features

adjectives,adj-verbs			nouns,un-known words		
Positive			Positive		
	明るい	0.62		満足	0.64
	広い	0.60		SXGA	0.57
	綺麗	0.58		インチ	0.56
	うれしい	0.57		買い物	0.56
	やすい	0.54		RAM	0.51
	快適	0.52		最高	0.51
	速い	0.50		買い	0.50
	軽い	0.50		GB	0.50
	細かい	0.49		メイン	0.50
	静か	0.49		利用	0.50
Negative	ひどい	-0.60	Negative	現象	-0.80
	駄目	-0.50		修理	-0.75
	異常	-0.49		症状	-0.73
	同様	-0.47		最悪	-0.68
	不安定	-0.47		連絡	-0.66
	ものすごい	-0.47		電源	-0.64
	正常	-0.46		サポート	-0.64
	いかが	-0.46		不良	-0.64
	真っ暗	-0.45		フリーズ	-0.64
	冷たい	-0.43		返品	-0.64

$$\begin{cases} \text{if } score(d) > 0 \rightarrow \text{positive} \\ score(d) < 0 \rightarrow \text{negative} \end{cases} \quad (3)$$

各文書に含まれる属性のスコアの総和が0より大きければ、肯定的な評判であると、0より小さければ、否定的な評判であるというように分類した。

3.3 文書分類に関する評価実験

3.3.1 PN分類器としての性能評価

分類の性能評価を行うため、比較対象として、C4.5およびSVMでも同様の実験を行った。C4.5は決定木学習のアルゴリズムの一つであり、情報利得に基づいて分類規則を学習する。また、SVMは近年その高精度・高速性を理由に注目されている。パーセプトロン型の二値分類問題に対する機械学習手法である。SVMにおいては、ツールとしてTinySVMを使用し、線形カーネルで実験を行った。他のオプションはデフォルトのままである。機械学習手法において与える属性については、スコアは用いずにその出現のみを考慮する形としたが、前章までで得られた属性と同様のものを用いた。訓練用のコーパスも同様に価格.comの2003年の肯定・否定の評判を用いた。

テストデータについては、価格.comの2004年²の使用レポート(良)・(悪)の書き込み、それぞれ、240件、137件を評判として利用した。各手法の分類精度については表2のようになった。

表2 PN分類の精度

Table 2 Accuracy of P/N classification

	(1)		(2)		(3)	
	P	N	P	N	P	N
Our approach	83.8	71.3	81.9	62.5	86.2	72.9
C4.5	79.1	60.5	78.6	58.2	78.0	60.3
SVM	79.6	71.8	81.7	66.2	80.4	73.0

本手法はC4.5よりP/N分類に関して確実に精度が高く、

² 正確には2004年4月8日までの書き込みを使用した

SVM と比較すると、Positive の分類精度は数%良い結果が得られ、Negative の分類精度は同程度であった。

また、分類の際に使用した属性の違いについては、Positive ではその差はあまり見られなかったが、Negative では形容詞、形容動詞を属性として採用するかどうかで差がでることが分かった。

3.3.2 スコアと分類精度の関係

評判らしい文書を抽出するフィルタとして、この分類手法を応用できないかを検討するためにテストデータにわざとノイズをいれ、高スコアが得られた文章が評判そのものとなることを理想的な結果と想定し、実験を行った。

価格.com の掲示板 2004 年の書き込みに対し、評判とは異なる文書³をノイズとして追加し、肯定的な評判、否定的な評判、ノイズを各 100 件になるようなデータセットを作成し、この実験でのテストデータとした。

この実験における結果は図 1 のようになった。

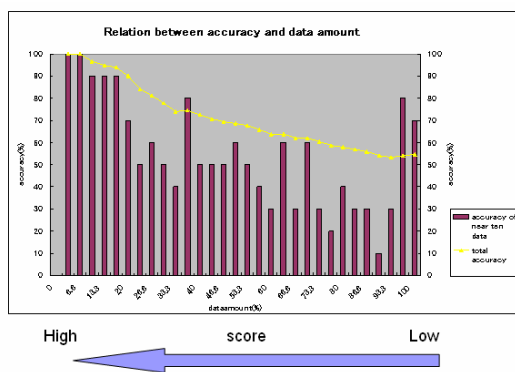


図 1 スコアと精度の関係

Fig.1 Relation between score and accuracy

まず、スコアの絶対値が大きい順にデータを並べかえ、10 個を単位としてそこまでのデータ全体の精度を求めたものが図の折れ線である。ただし、ここでの精度は肯定、否定、ノイズというように 3 値で分類した際の精度である。また、付け加えた直近の 10 個のデータの精度が図の棒グラフとなっている。図では、左から順に絶対値の大きい順にデータ量を増やしていき、一番右端では、折れ線はテストデータ 300 個全体での精度を表している。

この結果から、スコアが大きいものほど精度が高い、つまりスコアが大きいものは評判としても問題がないという結果が得られた。グラフの右端で直近 10 個のデータの精度が跳ね上がる傾向が見られる。これはノイズについては、評判でないというタグをつけて分類を行ったので、評判でないという意味で⁴精度が高かったためにこのような結果になっている。

4. Web からの評判抽出システムの試作

本章では、我々が試作した Web からの評判抽出システムに関する報告を行う。また、システムの試作によって新たに分かった問題、およびその考察について述べる。

4.1 システムの概要

³例えば、ノート PC の使い方の質問であったり、特価情報の噂など
⁴スコアが 0.0 となっている

前章までで述べた文書分類の手法を基にして、実際に Web から評判を抽出するシステムを試作した。システムの概要について以下に述べる。Crawling 部分については、今回は GoogleAPI を用い Google のデータベースを利用することとした。評判を検索する際には、利用者はノート PC のマシン名やその一部⁵をクエリとして入力する。クエリとして入力された文字列に「intitle:レビュー OR intitle:レポート」を付加したものを、Google へのクエリとして送信し、検索の結果得られた URL にアクセスし、HTML を入手する。入手したページのテキストを文単位で、PN 分類器にかける。得られたスコアの絶対値が上位の文から順に表示する。実際に、クエリ「VAIO V505」で評判を抽出した際の結果を図 2 に表示する。



図 2 評判抽出の例

Fig.2 An example of extracting reputations

図 2 の例では、対象の VAIO V505 のスコアの絶対値が大きい評判を抽出できているが、その中でも 4 番目の評判は VAIO typeS という製品のレビューの中に比較対象として V505 が登場していた。また、スコア下位では、やはりノイズが大量に含まれていた。

現段階での、抽出された結果の上位 5 件の精度⁶をクエリ 20 件⁷、計 100 件の抽出結果について調べたところ、精度 65% で評判であることが分かった。ただし、筆者が実際に抽出結果を読んで評判かどうかを判断したので現状ではまだ定性的な評価である。

実際に抽出に成功した例と失敗した例およびその原因に対して次頁図 3 に簡単に示す。

図 3 では現状のシステムでも評判が抽出される例を示したが、実際には評判抽出の精度や有効性を評価できるほどシステムはうまく機能していない。実用的な段階へシステムを高めていくには、問題点に対するさらなる検討が必要である。

4.2 現状での問題点とその考察

システムを試作した結果、実際に Web から評判を抽出する際には解決すべき問題点が大きく分けて 2 つあることが分か

⁵特に、型番を入れると良い結果が得られやすい
⁶ただし、ここではクエリに入れた検索対象の評判でなくとも、ノート PC の評判であれば正例として精度を算出している。
⁷価格.com のノート PC 人気アイテムランキングの上位 20 件の機種型の型番をクエリとした。ランキングのデータは 6/7 時点のものである。

<p>(成功例)</p> <ul style="list-style-type: none"> ・さらに重さ1.89kg、薄さも30mm弱と、「クリアブラック液晶」を搭載したワイドな画面とは裏腹に、軽く薄いのので携帯にはとても便利です。 ・USBコネクタの配置や筐体全般のデザイン、パッケージングなど、細かいところに不安や不満を感じる方も少なくはないかと思いますが、「持ち歩いて使うこと」を何よりも優先してデザインされ、まとめられたガジェットであることは疑いありません。 ・公称値で3.5時間の駆動が可能付属のACアダプタは本体に合わせて小型なものとなっている●薄さ、軽さのコストパフォーマンスという観点でX505を上回る本製品は以上のようにミニノートPCとしては十分なスペックを備えながら、重量は約910g、厚さは15.7mm(最薄部)と軽さ、薄さを実現している。 <p>(失敗例)</p> <ul style="list-style-type: none"> ・ここまでそうとう長かったですが、この先も十分に長いですから、カ・ク・ゴ・して読んでください(お※各画像をクリックすると大きい画像が表示されます。 - (原因: 評判とは全く無関係の話だが「大きい、長い、十分」という高スコアの属性を多数含む) ・便利なおすすめソフトをスペシャルプライスで大放出! - (原因: ノートPCの話題ではなく、単なる宣伝等の素晴らしいさを訴える表現を評判として抽出) ・(長浜和也) 2004年5月18日の記事使いやすさを重視した510万画素デジカメソニーDSC-W1ソニーのDSC-W1は、使いやすさを重視した510画素CCDを搭載したデジタルカメラである。 - (原因: 評判ではあるが対象がデジタルカメラであった)
--

図3 成功例と失敗例

Fig.3 An example of success/failure cases

った。以下、その問題点およびその考察について述べる。

まず、ひとつめが評判を含んでいるWebページの発見である。現状のシステムでは、抽出できる評判の量が非常に限られている。これは、Googleに送信するクエリの形に起因しているが、評判の量を増やすためにいたずらにGoogleで検索する範囲を広げると全く評判を含まないページの割合も大幅に増大し、結果として精度の悪化や処理時間の増大を招いてしまう。精度の向上のためには、リンク解析やテキストの類似性に着目したクラスタリング等によって、ある程度評判を含んでいそうなWebページ群を特定しておく必要があると考えられる。また、その特性上、評判が多く含まれていると考えられる掲示板から、現状ではほとんど抽出することができていない。価格.comや2ちゃんねる、Yahoo!掲示板など有名掲示板については、別モジュール等によって対応する必要があると考えられる。

次に、サイト内のHTMLの構造に起因する問題がある。この問題は2種類に分割されるのだが、1つは、Webページ内でたくさんの製品に対する評判が述べられていて、対象の製品以外の評判までも混同して抽出する問題である。もう1つは、そもそもそのWebページ内で対象の製品の評判は述べられていないがたまたま、テキスト内にその製品名が登場してしまう場合である。両問題において、フレームの構造や<table>タグや<p>タグ等にある程度の規則性が見られる場合が多いので、HTMLの構造を解析することによって精度の向上が期待できる。しかし、後者の問題では、例えば違う機種の評判の中で比較対象として対象の製品名が述べられている場合など製品名が登場する位置によっては、ノイズを除去するのが非常に困難であることが予想される。

今回報告した手法では、評判を抽出した際、その評判が本当に目的の対象の評判かどうかまで考慮されている手法ではなかった。今後、より実用的なシステムにしていくためには、この点を新たに考慮していく必要がある。

また、現在は文単位で評判の抽出を行っているが、上記のHTMLの構造解析とともに、テキストのセグメント単位での切り出しを行うことによって精度の向上が期待される。

5. まとめと今後の課題

本報告では、日本語での評判の P/N 分類について、統計

的な処理を用いた手法について実装、評価実験を行った。本手法は従来から用いられてきた機械学習手法と比較してほぼ同程度の精度が得られることが分かった。また、高スコアの文書は評判そのものであることも確認し、Web 上から評判のような文書を抽出する分類器として応用できる可能性があることを示した。最後に、実際に Web から評判を抽出するシステムを試作し、抽出を行った例について示した。また、システムの構築の結果生じてきた問題に対し考察を行った。

今後は、本報告での試作システムの改良、および他ドメインへの拡張等を行っていく予定である。

【文献】

- [1] 池尾恭一 編. ネット・コミュニティのマーケティング戦略,有斐閣,2003.
- [2] 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, NL-144-11, pp.75-82, 2001.
- [3] Kushal Dave, Steve Lawrence, David M.Pennock. Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews. International World Wide Web Conference(WWW2003) pp.519-528, 2003
- [4] 小林のぞみ, 乾健太郎, 松本裕治, 立石健二, 福島俊一. テキストマイニングによる評価表現の収集. 研究報告「自然言語処理」No.154, 2003
- [5] Bo Pang, Lillian Lee, Shivakumar Vaithyanathan. Thumbs up? Sentiment Classification using Machine Learning Techniques. Empirical Methods in Natural Language Processing(EMNLP2002) pp.76-86, 2002.
- [6] Hong Yu, Vasileios Hatzivassiloglou. Towards Answering Opinion Questions: Separating Facts from Opinions and Identifying the Polarity of Opinion Sentences. Empirical Methods in Natural Language Processing(EMNLP2003), 2003
- [7] 乾孝司, 乾健太郎, 松本裕治 : 出来事の望ましき判定を目的とした語彙知識獲得, 言語処理学会第 10 回年次大会発表論文集, 2004.3.

藤村 滋 Shigeru FUJIMURA

東京大学大学院情報理工学系研究科修士課程在学中. 2002 東京大学工学部電子情報工学科卒業. Web マイニングに関する研究に従事. 日本データベース学会学生会員.

豊田 正史 Masashi TOYODA

東京大学生産技術研究所特任助教授. 1999 東京工業大学情報理工学研究科博士後期過程修了, 理学博士. Web マイニング, ユーザインタフェース, ビジュアルプログラミングの研究に従事. 日本データベース学会正会員. 情報処理学会, 日本ソフトウェア科学会, ACM, IEEE CS 各会員.

喜連川 優 Masaru KITSUREGAWA

東京大学生産技術研究所教授. 2003 年 4 月より, 同所戦略情報融合国際研究センター長. 1983 東京大学大学院工学系研究科情報工学博士課程修了, 工学博士. データベース工学, 並列処理, Web マイニングに関する研究に従事. 本会理事, 情報処理学会理事・フェロー, SNIA-Japan 顧問, ACM SIGMOD Japan Chapter Chair(H11-H14), 電子情報通信学会データ工学研究専門委員会委員長(H9,10), VLDB Trustee, IEEE TKDE Assoc.Editor, IEEE ICDE, PAKDD, WAIMSteering Comm.Member.