社団法人 情報処理学会　研究報告
IPSJ SIG Technical Report

2004－DBS－134　(II)　(63)
2004／7／14

Webからの評判および評価表現抽出に関する一考察

藤村 滋 †　　豊田 正史 ‡　　喜連川 優 ‡

† 東京大学大学院情報理工学系研究科
‡ 東京大学生産技術研究所

要 旨

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．

A Consideration of Extracting Reputations and Evaluative Expressions from the Web

Shigeru FUJIMURA†　　Masashi TOYODA‡　　Masaru Kitsuregawa‡

†Graduate School of Information Science and Technology, The University of Tokyo
‡Institute of Industrial Science,The University of Tokyo

Abstract

Automatic extraction of reputations from huge information of the Web is much attentioned. Extracted reputations can be utilized for marketing, claim management in companies and decision support in customers. But, because of subjective aspects of repuattions, 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るのは明白である．[2]

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．肯定・否定への分類のように主観的な情報を基にした文書分類では，注目すべき属性が，従来から行われてきたトピックを基にした文書分類とは異なってくる．そこで，我々はあらかじめ収集しておいた肯定・否定の評判から評価表現を抽出し，属性とした．また，分類器が評判の良し悪しを決めるルールそのものとなるが，機械学習による分類器では人間にとって解釈しずらいという問題点があったため，解釈可能な分類器を作成し，その精度を調べたところ機械学習手法と同程度であることが得られた．さらに，提案手法を Web からの評判の抽出に応用するための実験を行い，評判抽出のシステムを構築した．その結果，実際に評判抽出のシステム構築に当たって，様々な問題が生じることが分かった．

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る．そして，4 章では評価実験について述べ，5 章でその考察・検討について述べる．次に，6 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 7 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 [6] 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，適正値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[1] 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．

この他に評判に関する研究や出来事の望ましさという尺度で文書分類や知識獲得を行うといった研究の例としては，以下にあげる研究がある．

まず，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[7] らの研究もある．さらに，この分野において最近発表された論文の例としては，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 [3] の研究があげられる．

3 本報告の手法

本章では，まず全体の処理の流れについて述べ，次に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る．

3.1 全体の処理の流れ

我々の最終的な目標は，Web 全体からの評判抽出を行うシステムを構築することにある．最終目標となるシステムに必要な機能および処理の流れについては，図 1 に示す．そこで，まず評判が抽出できたと仮定して，その評判の肯定・否定分類[1]に取り組むこととした．

[1]以下，PN 分類と呼ぶ

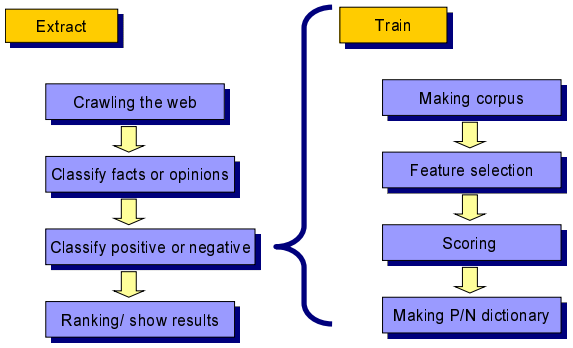

図 1: Process flowchart of our approach

本手法の PN 分類部分では，次のことに注意した．まず，ドメインに限定されることがないようにする点である．次に分類器を容易に解析することができるという点である．後者については，評判を PN に分類する分類器は，評判の良し悪しを決定するルールそのものであるから，その分析を行うことによって，現在のトレンドや潜在的なニーズを掴むことができる可能性がある．

以上の要件を満たした PN 分類器作成のために，統計的に評価表現を取り出すことでドメイン依存の問題を解決した．コーパスから属性を選択し，評価表現としての重みをスコアリングし，評価表現辞書を作成する処理を行った．

3.2 評価表現辞書の構築

- **評価表現およびその辞書**

  本報告でどのような意味で評価表現・評価表現辞書という言葉を用いるかについて次のように説明する.「評価表現とは，評判で用いられる特徴的な“ 語 ”のことであり，評価表現辞書とはその評価表現が肯定・否定どちらの表現であるかまで記された評価表現の集合である.」

- **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，茶筅[2]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げられると考えている．

- **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．

$$score(w_i) = \frac{P_P(w_i) - P_N(w_i)}{P_P(w_i) + P_N(w_i) + k}$$
$$(-1 \leq score(w_i) \leq 1) \qquad (1)$$

[2] http://chasen.aist-nara.ac.jp/hiki/ChaSen/

表 1: Hi-scored features

adjectives,adj-verbs			nouns,un-known words		
Positive	明るい	0.62	Positive	満足	0.64
	広い	0.60		SXGA	0.58
	綺麗	0.58		インチ	0.57
	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

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．

3.3 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．

$$Score(d) = \sum_{ALL w_i} score(w_i) \tag{2}$$

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

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し，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．

4 評価実験

本章では，PN 分類器としての性能を評価するために行った実験，および評判を含んだ意見と事実を分類する分類器への応用の可能性を調べるために行った実験について報告する．

表 2: Accuracy of P/N classification

	(1)		(2)		(3)	
	P	N	P	N	P	N
Our approach	83.8	71.3	84.9	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

4.1 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[3]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[4]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[5]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 3 のようになった．本手法は C4.5 より P/N 分類に関して確実に精度が高く，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] http://chasen.org/~taku/software/TinySVM/
[4] http://www.kakaku.com/
[5] 正確には 2004 年 4 月 8 日までの書き込みを使用した．

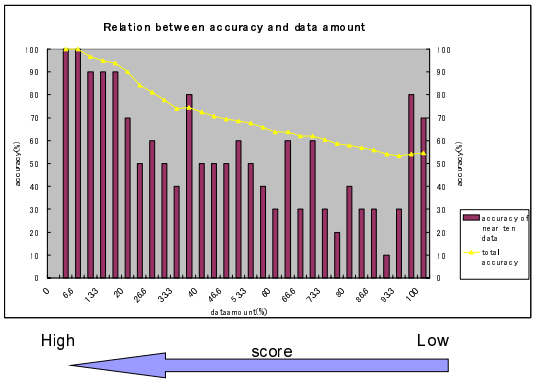

図 2: Relation between score and accuracy

4.2 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[6]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 2 のようになった．

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[7]精度が高かったためにこのような結果になっている．

[6]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

[7]スコアが 0.0 となっている

5 考察・検討

本章では，実験の結果得られた精度についての考察を行う．また，評価表現辞書の分析によってさらなる知識獲得の可能性があること，および，評価表現辞書における属性に関する考察について述べる．

5.1 本手法の精度について

今回の実験の 1 つの目的であった Dave らの手法を参考にした手法が，日本語においてどの程度有効であるかを調べるという点については，英語の場合では精度 85%程度であったのに対し，日本語ではおよそ精度 70%代後半であった．結果として精度については日本語のほうが 6%前後劣っていた．さらに，この実験の結果を得る前に，予備実験を行った際，日本語では品詞による選別を行わないと精度が大幅に低くなることが分かった．英語では特にストップワードを設けるようなことをしなくとも問題がなかったのに対し，日本語では形容詞，形容動詞，名詞，未知語のように，評価表現としての属性を絞る必要性があることが分かった．

また，機械学習による分類との精度比較では C4.5 よりは良い結果が得られたが，SVM との差はそれほど大きくなかった．しかし，本手法には分類器の分析による知識の獲得という大きなメリットが存在する．SVM の分類器は人間にとって可読不可能であり，なぜ入力文書が肯定的な評判であるのかは人間には理解不能である．また，C4.5 の決定木は人間にとって可読性はあるが，ある語が出現するかしないかだけの 1 方向の決定木になる傾向が見られ，結果として得られる知識は少ない．

実際の知識獲得の例については，次節で述べることとする．

5.2 P/N 評価表現辞書分析による知識獲得

この節では，分類器の一部である評価表現辞書を分析することで得られた知識について述べる．

表 1,2 に示した属性の例に着目すると，「明るい」「綺麗」「SXGA」「インチ」などノート PC では主

に液晶について述べる際に用いられる語が高いスコアをもっている．ここから，ノート PC を購入する際には液晶に対する注目度が高いと推測される．また,「電源」という語自体がかなり否定的な評価表現であることは直感的には理解できない．しかし，このスコアは「電源」という語を用いた文書は大半が否定的な評判であるという事実を示す．

以上 2 つの推測を基に，コーパスの文書を実際に読んでみたところ，確かにノート PC の液晶に注目している人が多く，特に，ツルツルしたフィルムの様なものを張った液晶に対して注目度が高いことが得られた．また,「電源」については「電源が壊れる・故障する・入らない」といった類の文書が多く，確かに否定的な評判が多いことが得られた．

以上から，本手法では評価表現辞書を分析することによって，評判から一歩踏み込んださらなる知識を獲得できる可能性があることが示された．

5.3 精度とスコアの関係を調べる実験の際における属性について

精度とスコアの関係を確かめた実験の際，用いた属性は (1) 形容詞，形容動詞であった．(3) (1) ＋名詞，未知語の場合についても実験を行ったが，精度はデータ量が増えても振動するのみで期待された結果は全く得られなかった．

未知語は大半が名詞である．したがって，名詞を属性に加えることで実験結果が変わることとなる．この原因については，次のように解釈ができる．

前節で述べた「電源」の例について，確かに否定的な評判の中で使われる可能性が高い語であることは間違いない．しかし，例えば,「電源まわりがすばらしい」のように，肯定的な評判で使われる可能性も否定できない．つまり，形容詞，形容動詞に比べて，名詞は使われ方によって評判の良し悪しが変わってしまう可能性が大きい．評価表現辞書が統計的に作られているといっても，総計数十万単語からなるコーパスの中で多くても数十回程度の頻度の属性がほとんどであるため，名詞の使われ方によって評判の良し悪しが変わる可能性は統計的な手法でも吸収しきれないほどに大きいと推測される．名詞を属性として採用する際には，形容詞よりも影響を弱くするように実数 $\alpha(\alpha < 1)$ をかけるなどの工夫を検討する必要がある．

6 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．

6.1 システムの概要

前節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[8] 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[9]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 3 に表示する．

図 3 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[10]をクエリ 20 件[11]，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図 4 で簡単に述べる．

以上，図 3 では現状のシステムでも評判が抽出さ

[8] http://www.google.com/apis/

[9] 特に，型番を入れると良い結果が得られやすい

[10] 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．

[11] 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．

図 3: An example of extracting reputations

（成功例）
・さらに重さ 1.89k g、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。

（失敗例）
・ここまでも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（ぉ※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 4: An example of success/failure cases

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．

6.2 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性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[12]，Yahoo! 掲示板[13]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性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．

[12] http://www.2ch.net/

[13] http://messages.yahoo.co.jp/index.html

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．

7 おわりに

本報告では，日本語での評判の P/N 分類について，知識獲得が容易になるように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してくるような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．

以下，今後の課題について列挙する．

- **評判抽出システムの改良**
  前章の考察で述べたことを中心に改良を行っていく予定だが，まずは HTML の構造解析から取り組むこととする．その後は，最も重要な課題である Web 上での評判を含む可能性の高いページ群の特定に取り組んで生きたい．

- **他ドメインへの拡張**
  本報告では，ノート PC に対象を絞って実験を行ったが，今後はデジカメや液晶ＴＶなどの他デジタル家電やレストラン，映画といったドメインの評判抽出にも拡張していく．

- **コーパス量と精度の関係の検討**
  評判のコーパスを作成したり，コーパスとして使えそうな文書を Web などから発見し，利用できる形に変換するのは容易な作業ではない．そこで，どの程度のコーパスの量があれば十分なのかを確認するために，コーパスの量と精度の関係について検討していく．

参考文献

[1] 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

[2] 池尾恭一 編. ネット・コミュニティーのマーケティング戦略, 有斐閣,2003

[3] 乾孝司, 乾健太郎, 松本裕治：出来事の望ましさ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.

[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 Clasification using Machine Learning Techniques. Empirical Methods in Natural Language Processing(EMNLP2002)pp.76-86,2002.

[6] 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.

[7] Hong Yu,Vasleios Hatzivassiloglou. Towards Answering Opinion Questions:Separating Facts from Opinions and Identifying the Polarity of Opinion Sentences. Empirical Methods in Natural Language Processing(EMNLP2003),2003