

文の構造を考慮した評判抽出手法

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 〒 113-0033 東京都文京区本郷 7-3-1

^{††} 東京大学生産技術研究所 〒 153-8505 東京都目黒区駒場 4-6-1

E-mail: †{fujimura,toyoda,kitsure}@tkl.iis.u-tokyo.ac.jp

あらまし Web 上の評判を扱うためには、評判を構成し、その理由となるような評価表現を特定することが重要である。従来の研究では、評価表現として形容詞・形容動詞等の単語を用いたり、文の構造をあらかじめ規定した上で素性を検討することが多かった。しかし、評判の理由を知るには単語では不十分であり、文の構造を恣意的に用いた場合では、規定した文構造以外の評判を扱うことが不可能になるという問題があった。そこで、本報告では文を構成する上で主要な単語のみを用い、文節の n-gram による素性を提案する。また、提案する素性を用いた、評判の肯定・否定分類法を提案・評価する。最後に、評判抽出への応用について述べる。

キーワード 評判, テキスト分類, テキストマイニング

A Reputation Extracting Method Considering Structure of Sentence

Shigeru FUJIMURA[†], Masashi TOYODA^{††}, and Masaru KITSUREGAWA^{††}

[†] Graduate School of Information Science and Technology, The University of Tokyo

7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-0033, Japan

^{††} Institute of Industrial Science, The University of Tokyo

4-6-1 Komaba, Meguro-ku, Tokyo, 153-8505, Japan

E-mail: †{fujimura,toyoda,kitsure}@tkl.iis.u-tokyo.ac.jp

Abstract It is important to specify evaluation expression which is a reason of reputations so as to deal with reputations on the Web. In recent researches, features which are made by words of adjectives and adjective verbs are often adopted. However, word-level features are inadequate to understand the reason of reputations. In other way, features using arbitrary structure of sentence are often adopted too. But these features cause decrease of coverage. In this paper, we propose features which are made by only main words in meaning of sentences. We also propose the method of classifying reputations using proposed features and evaluate this method. At the end, we describe application of classifier to extracting reputations.

Key words Reputation, Text Classification, Text Mining

1. はじめに

Web の爆発的な成長および急速な普及と共に、個人の情報発信の場としての Web への注目が集まっている。個人が発信する情報の中でも、とりわけ意見や評判といった主観的な情報が重要視されている。主観的な情報を自動的に収集・抽出・分類することで、個人においては商品やレストラン等の興味対象への他人の評価の収集支援や企業においてはマーケティングやクレーム処理の支援への応用が期待されている。

主観的な情報の中でも、特に評判を扱う際には、肯定的か否定的かに分類されていることがその後の分析を容易にするという意味で望ましい。したがって、テキスト分類技術の応用が期

待される。従来のテキスト分類に関する研究では、例えば新聞記事を政治、経済、スポーツ等のカテゴリに分類するといったトピック中心の分類が話題となってきた。このような、トピック中心の分類では一般に単語、特に名詞の集合を素性とすることが多く、このような単純な素性でも十分な精度が達成されてきた。一方で、評判の肯定・否定分類のような主観的な情報に基づいた分類では、従来のトピック中心の分類で用いられてきた素性では十分な精度は得られにくく、肯定・否定の評判の原因となるような評価表現を素性にしなければならぬ。

従来の評判を扱う研究では、評価表現として形容詞や形容動詞等の単語や文の構造を恣意的に用いた素性を採用することが多かった。日本語では対象への評価を表す際には、主に形容詞

や形容動詞が用いられる。しかし、形容詞、形容動詞には、例えば「良い」「悪い」のように絶対的な評価を与える語もあれば、「短い」のように周囲の語や文脈によって、評価が変わってくる語もある^(注1)。したがって、単語レベルの素性では評判の理由としては不十分である。また、例えば、「～(評価の対象)(が)は)～(評価語)」のような構造を持つ文を評判とするというように、文の構造を恣意的に用いた素性の場合では、その規定した文構造以外の評判を扱うことができないという欠点があった。

そこで、本稿では文を構成する主要な語のみを用いた、文節 n-gram を素性として採用する。n-gram であれば、語の出現順を考慮する形となり、さらに、主要な語のみを用いる事で文の構造を維持することを狙いとしている。この素性を用いて、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、肯定・否定の文書分類を応用した評判抽出を行い、その評価および考察について報告する。

以下、2章では関連研究について述べる。3章では提案手法についての説明を行う。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、および事例研究について述べ、最後に7章で本報告のまとめと今後の課題について記す。

2. 関連研究

評判の抽出に関する先行研究としては、立石 [9][10] らの研究があげられる。この研究では、ユーザが入力したクエリ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、評価表現は話題のドメインによって大幅に変わり、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いという問題点があった。

そこで、評価表現や主観的な表現を自動的に収集する研究が [2][3][4][8][11] のように行われてきている。

工藤ら [5] は評価表現収集が研究の目的ではないものの、主観的な文書分類に向けて、従来の単語レベルの素性ではなく、係り受け木の部分木を素性とすることでより意味的なまとまりを素性とすることを提案した。ただし、助詞なども素性の一部として採用しているという点が精度に悪影響を与えていると考えられる。

一方、Web 上のレビューを肯定・否定に分類する研究の例としては、Dave [1] らの研究がある。この研究では、一般的な語は肯定、否定の評判にも同程度に出現するはずであるから、肯定と否定の評判の差をとることによって評判を特徴付けるような表現を獲得することができ、その表現を含む文書を肯定・否定の評判とするという手法を提案している。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 [6] 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 [12] らの研究もある。

意見や評判、口コミなどの主観的な情報を検索出来るツールとして実際に、公開・販売されているものとしては、立石 [9] らによる「eHyouban」や、鈴木ら [8] も関わっている「blogWatcher」、相良ら [7] による「レストランのウワササーチ」などがある。ただし、相良らは直接主観的な情報を検索することを目的としているわけではなく、レストランにおける場所という情報に着目して Web 上の情報を収集している。

3. 提案手法

本章では、まず本稿で評判を扱うにあたり、いくつかの準備段階としての考察および説明を行う。次に、提案する素性についての説明を行い、素性を肯定的な表現か否定的な表現かに区別するためのスコアリングの方法について説明を行う。最後に、評判の肯定・否定分類を基にした評判抽出法についての提案を行う。

3.1 準備

本報告で実際に取り扱うドメインとしては「ノート PC」とし、評判を扱う際の粒度については、文単位とする。より大きな粒度で評判を扱った場合には、文単位の場合と比較した際に、肯定・否定両方の評価を含む事例が多く見うけられ、結果として分類が曖昧になるという問題が生じるためである。

文単位で評判を扱うためには、Web 上のテキストを文単位に分割する必要がある。文分割の手法としては、まず句読点「。」や、疑問符「？」や感嘆符「！」等、文の区切りとなりうる記号や HTML タグを利用して文単位に区切った。また、肯定・否定のあいまい性をなくす為、逆説の接続助詞「が」で文を区切る等、簡単なヒューリスティクスを構築し文分割を行っている。

また、実装においては形態素解析ツールとして「茶筌」^(注2)および文節に区切るために用いた係り受け解析器としては「CaboCha」^(注3)を用いた。

最後に、評価表現辞書を作成するための訓練用コーパスとして利用するため、価格.com のノート PC に関する掲示板の 2003 年 9 月～2004 年 12 月の書き込みを人手で肯定・否定の評判に分類を行った。肯定の評判 3148 文、否定の評判 2668 文を訓練用コーパスとした。

また、実験の際にはこのコーパスを用いて、5 分割交差検定を行うことによって評価を行っている。

3.2 本報告で採用した素性

藤村 [2] では、形容詞、形容動詞、名詞、未知語といった単語レベルの素性を用いたが、次のような課題があった。まず、形容詞「大きい」のように、例えば「画面が大きい(肯定的)」、「騒音が大きい(否定的)」といった、前後の語によって、肯定・否定が変化する場合に対応できなかった。次に、名詞「電源」は否定的な語であったが、直感的にはなぜ「電源」が否定的な語なのか分からない。しかし、実際にこれらの語を含む文書を調べると、ほとんどが、「故障して電源が入らなくなった」と

(注1): 一般的に「処理時間が短い」ならば肯定的、「バッテリー寿命が短い」ならば否定的な表現であろう。

(注2): <http://chasen.aist-nara.ac.jp/hiki/ChaSen/>

(注3): <http://chasen.org/taku/software/cabocho/>

例)このマシンのCPUファンの音がうるさすぎます。

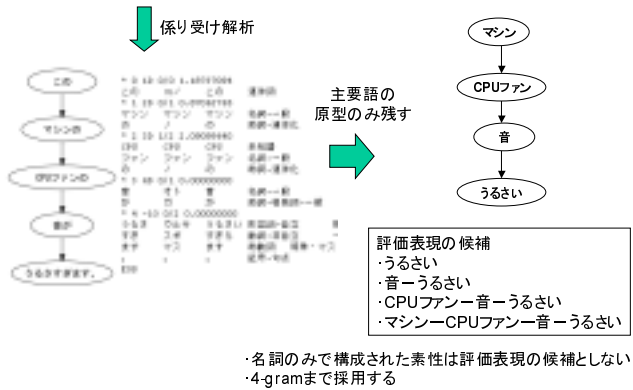


図1 素性の決定の仕方

といった旨の否定的な書き込みであることが分かった。以上より、確かに肯定・否定のどちらかに偏って現れやすい名詞は存在するが、評判の分析という観点からすると、名詞単独では不十分で周囲の情報が必要となる。

以上の課題を解決するため、本報告では、文を構成する主要な単語のみを用いた文節 n-gram を素性として採用した。文節 n-gram では語順が保持された状態で、共起性の高い連続した句が素性として採用される。この共起性の考慮により、周囲の語によって肯定的か否定的かが変わる語への対応を図ることが期待される。また、語の連なりが長くなるほど評判の分析も容易になると考えられる。

素性の詳細については以下で述べる。また、実際の処理の様子は図1のようになる。以降の素性の例などでは、単語の基本形のみしか考慮していないような表記の形態をとるが、実際には、内部では基本形と品詞の情報の2つを保持して処理を行った。

- 一文の係り受け解析結果から、主要な語のみを残してそれ以外の語は取り除く
- 残す語としては、名詞^(注4)、未知語^(注5)、動詞^(注6)、形容詞、および否定の助動詞「ない」^(注7)
- n-gram^(注8)統計をとり一定の出現頻度かつ用言を含んでいるものを素性として採用する

一方で、より文構造を考慮するには、係り受けの情報も用いて、係り受け木の部分木を素性として採用する手法も考えられる。

この点については、予備実験から、次のような知見を得た。係り受け木の部分木を素性として使った場合のメリットとして、例えば「画面がとても大きい」といったような、副詞を含んだ文を「画面が大きい」とマッチさせることができるという点が

ある。その一方、大半の素性については、係り受け木を用いた場合でも、文節 n-gram を用いた場合でも同一であることが分かった。よって、本報告では副詞を取り除くことにより、上記の副詞に起因する問題を回避した。また、現状では係り受けの情報は文節への区切りと比較し誤っていることも多く、ボトルネックとなってしまう可能性もある。この原因としては、対象となる Web 上のテキストが口語ドメインであり、また、家電や PC といった未知語を多く含むドメインのため新聞をコーパスとして学習する係り受け解析器には非常に困難なタスクとなっていることが考えられる。

3.3 否定の表現に関する処理

助動詞「ない」^(注9)は、直前の動詞・形容詞等の意味を反転させる語である。そこで、文中に出現する助動詞「ない」の個数が偶数個の場合は素性のスコアをそのまま用い、奇数個の場合は素性のスコアを反転させるという処理を行った。

また、「～ません」「～ませんでした」といった表現も否定の表現である。詳しくは次節で述べるが、スコアリングを行うための訓練コーパス中における出現回数を調べたところ、出現頻度が 446 回であり、無視すると評価表現辞書において精度の悪化が懸念されることが分かった。そこで、本報告では簡単なヒューリスティクスにより、「～ない」という表現に置き換えることで対処することとした。

3.4 評価表現のスコアリングとその信頼性の検討

肯定の評判での出現頻度と否定での出現頻度の差分を取ると、一般的な意味で使われる素性は打ち消しあい絶対値が 0 に近いと考えられる。一方、例えば肯定的な表現は肯定での出現頻度のほうが大きいと考えられるので正の値を持つと考えられる。この、仮定に基づき評価表現のスコアリングを行った。実際には次の式のようなになる。

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

$$(-1 \leq score(w_i) \leq 1) \quad (1)$$

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 に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。

一方で、肯定・否定の評判が与えられたとき、肯定(否定)の評判のみに偏って現れる素性は肯定(否定)の評価表現であると考えられる。評価表現でない素性は、与えられた肯定の評判、否定の評判の比率通りに現れると考えられる。そこで、 χ^2 値を用いることによって、その素性がどれだけ偏って現れるかを数値化し、 χ^2 検定によって信頼性による素性の選別を検討する。

ここでの χ^2 値は次のような式で求められる。

(注4): 代名詞、非自立語を除く。一方、茶筌の結果では形容動詞は名詞・形容詞語幹として名詞に含まれる。

(注5): ただし、名詞として。

(注6): 非自立語を除く

(注7): 否定因子として別処理を行う。

(注8): 実際にはコーパス量を考慮し、four-gram までとした。

(注9): 茶筌においては「助動詞 特殊・ナイ」として扱われる

$$\chi^2(w_i) = y_{w_i} \cdot \sum_{w, class} \frac{(df(w, class) - df_{total} \cdot p_w \cdot p_{class})^2}{df_{total} \cdot p_w \cdot p_{class}} \quad (2)$$

$$w \in \{w_i, \bar{w}_i\} \quad (3)$$

$$class \in \{positive, negative\} \quad (4)$$

$$\begin{cases} \text{if } p_{w_i} > p_{\bar{w}_i} \rightarrow y_{w_i} = 1 \\ p_{w_i} < p_{\bar{w}_i} \rightarrow y_{w_i} = -1 \end{cases} \quad (5)$$

ただし, $df(w, class)$ は肯定 (否定) の評判で, 素性 w_i を含む (含まない) 文の頻度であり, df_{total} は全文書数, p_w は素性 w_i が生起する (しない) 確率, p_{class} は全文書にしろ肯定 (否定) の評判の割合である.

最後に, 実際に作成された評価表現辞書については巻末に Appendix として示す.

3.5 肯定・否定分類法と評判抽出への応用

藤村 [2] では, 評判の肯定・否定分類を応用することによって, 評判抽出を肯定・否定の評判, ノイズという 3 値分類問題に置き換えることを検討している. 本報告では上記の手法を応用して評判抽出を行う.

まず, 一文が未知のデータとして与えられたとき, その文に対するスコアリングが必要となる. スコアリング法については次の式で示す

$$Score(sentence) = \sum_{ALL w_i} score(w_i) \quad (6)$$

$$\text{ただし, } w_i \notin w_{parent} \quad (7)$$

$$\begin{cases} \text{if } Score(sentence) > 0 \rightarrow positive \\ Score(sentence) < 0 \rightarrow negative \end{cases} \quad (8)$$

入力文中に存在する素性のスコアの総和を入力文のスコアとするのだが, n-gram 素性にしたことによって, 式 (7) の条件が重要になる. もし trigram の素性「ドット抜ける-ない-いい」が十分な出現頻度を持って, 評価表現として採用されている場合, 結果としてこの素性の一部からなる, 例えば「ドット抜ける-ない」や「ない-いい」という素性についても評価表現として採用されてしまう. よって, 入力文のスコアリングには親となる素性を持たない素性のみが必要となる.

入力文のスコアリングの結果, スコアが正でその絶対値が大きいほど強い肯定の評判であり, 逆にスコアが負でその絶対値が大きいほど強い否定の評判ということになる.

4. 評価実験

本章では, 精度・再現率により, χ^2 検定による素性の選別, 評判分類法として n-gram 素性の効果, および機械学習法との比較について実験し評価を行ったので報告する.

4.1 χ^2 検定による素性の選択

χ^2 値は偏りの優位さつまり, どの程度の信頼性を持って偏っているかを示す数値である. そこで, χ^2 検定を行って, 信頼性

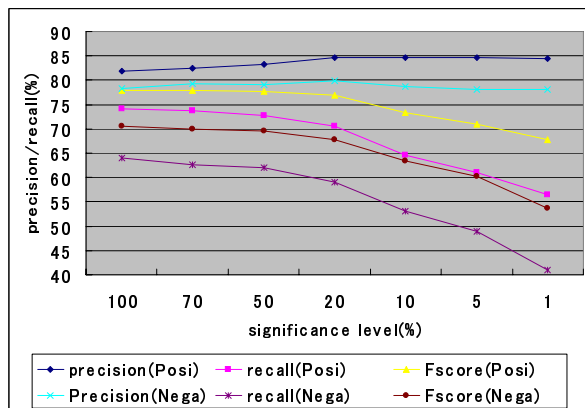


図 2 肯定・否定の精度, 再現率と優位水準

表 1 素性数と優位水準

significance level(%)	100	70	50	20	10	5	1
feature(s)	1928	1568	1465	1210	407	245	118

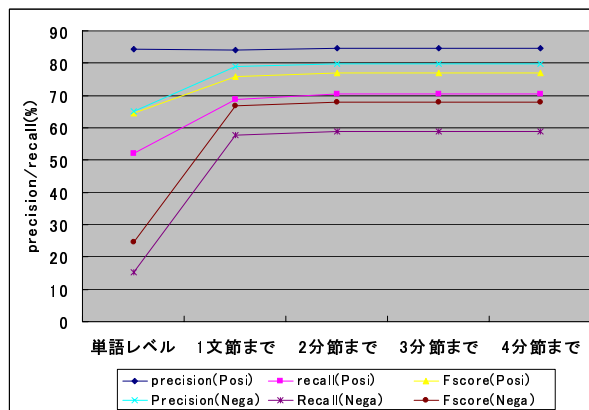


図 3 肯定・否定について素性の違いと精度, 再現率

表 2 素性数

	単語	1文節	2文節まで	3文節まで	4文節まで
feature(s)	423	580	1123	1193	1210

の低い素性を却下し, 精度・再現率が最高となる優位水準を求めるための実験を行った. その精度・再現率および F 値については図 2, 表 1 に示す.

結果として, 優位水準が 50% までは, 素性数の減少によって再現率が低下し, 結果として F 値が下がることはなかった.

したがって, 今回の報告では優位水準 50% を採用し, 今後の実験を行うこととする.

4.2 文節単位での処理のメリット, および n-gram 素性の効果

文節単位での処理のメリットを示すため, 単語レベルの素性との比較実験を行った. 単語レベルの素性については, 形容詞および名詞 (形容動詞-語幹とナイ形容詞-語幹) を素性とした. 一般的な名詞, および動詞を素性としなかったのは, 評判の理由, つまり特徴量とみなすには不十分な語が多く. 例え, 精度や再現率が高かったとしても本当の意味での評判分類は行うことができていないと考えられるためである.

表3 機械学習法との比較

	Our approach		C4.5		SVM	
	Precision	Recall	Precision	Recall	Precision	Recall
Positive	84.7	70.5	82.2	65.5	84.2	74.8
Negative	70.3	84.5	66.6	82.8	73.1	83.1

文節単位の素性と単語単位の素性について、精度・再現率は図3、表2で示される。

結果として、単語レベルの素性から文節レベルの素性になることで大幅な再現率の向上^(注10)が得られることが分かった。一方 n-gram 素性については、bi-gram までは精度・再現率の向上に貢献することが分かったが、trigram 以上では精度の貢献は得られなかった。

4.3 機械学習による分類との比較

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^(注11)を使用し、多項式カーネルで実験を行った。他のオプションについては、精度・再現率の結果から決定したところ、結果的に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。訓練用コーパスおよび評価法についても同様である。

また、SVM と C4.5 においては実装上の問題として素性が現れなかった場合、すべて否定の評判としてシステムは回答を出すようになっていた。提案手法は、素性が存在しなかった場合、分類不能と言う出力を返すようにしているが、ここでは公平な比較を行うため、我々も全て否定として出力するようにした。 χ^2 値の式より、結果として肯定の精度以外に何らかの影響を与えるものと考えられる。

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、肯定の再現率および、否定の精度について 3~4% 程度劣っている。

しかし、本手法には SVM にはない、分類器、有効な素性の解析というメリットがある。

5. 考察・検討

本章では、まず、評価表現辞書の分析に関する考察を行う。次に、抽出が失敗する原因についての考察を行う。

5.1 評価表現辞書の分析

提案手法の利点として、評価表現辞書の分析によって知識の獲得に役立つことが上げられる。以下、実例を基に考察を行う。

● 「ドット」を含む素性

「ドット」を含む素性について、表4に示す。

表4 名詞「ドット」を含む素性

	頻度 (肯定)	頻度 (否定)	スコア	χ^2 乗値
ドット抜ける-ない	11	1	0.802	6.64
ドット欠け-ない	4	0	0.768	3.318
液晶-ドット抜ける-ない	3	0	0.713	2.49
ドット抜ける-ない-いい感じ	2	0	0.623	1.658
ドット抜ける-はなし	2	0	0.623	1.658
ドット抜ける-ある	0	3	-0.783	3.62

表5 形容詞「大きい」を含む素性

	頻度 (肯定)	頻度 (否定)	スコア	χ^2 乗値
大きい	30	15	0.247	2.633
液晶-大きい	3	0	0.713	2.49
大きい-綺麗	2	0	0.623	1.658
画面-大きい	2	0	0.623	1.658
文字-大きい	2	0	0.623	1.658
大きい-見る	2	0	0.623	1.658
大きい-見る-すぐ	2	0	0.623	1.658
動作音-大きい	0	2	-0.707	-2.413
音-大きい	0	2	-0.707	-2.413
大きいなる	0	3	-0.783	-3.62
大きいなる	0	3	-0.783	-3.62
音-大きいさ	0	3	-0.783	-3.62

表6 名詞「起動」を含む素性

	頻度 (肯定)	頻度 (否定)	スコア	χ^2 乗値
起動-早い	2	0	0.623	1.658
起動する-液晶	0	2	-0.707	-2.413
アプリ-起動する	0	2	-0.707	-2.413
起動する-液晶-暗い	0	2	-0.707	-2.413
起動-かかる	0	2	-0.707	-2.413
起動-する	0	3	-0.783	-3.62
起動する	14	36	-0.513	-14.6

表7 名詞「キーボード」を含む素性

	頻度 (肯定)	頻度 (否定)	スコア	χ^2 乗値
キーボード-打つや-すい	3	0	0.713	2.488
言う-キーボード	3	0	0.713	2.488
キーボード-打つや	3	0	0.713	2.488
キーボード-いう	2	0	0.623	1.658
キーボード-打ちやすい	2	0	0.623	1.658
キーボード-タッチ-良い	2	0	0.623	1.658
キーボード-打ち-やすい	2	0	0.623	1.658
キーボード-打つ	0	2	-0.706	2.413
キーボード-打つづらいさ	0	2	-0.706	2.413
キーボード-コーティング-剥げる	0	2	-0.70691	2.413

液晶画面に対する話題の際、液晶のドット抜けの有無を気にする人が多い。もちろん、液晶のドット抜けがないほうが望ましいことは言うまでもない。「ドット抜ける-ない」「ドット抜ける-ある」がそれぞれ、肯定・否定の評価表現でありドメイン知識をよく反映していると考えられる。

● 「大きい」を含む素性

「大きい」を含む素性について、表5に示す。

既に、述べたように「大きい」「小さい」「長い」などは、対象となるものによって肯定的か否定的かが変化する形容詞である。

実際に、「液晶-大きい」「画面-大きい」「文字-大きい」が肯定的、「音-大きい」「動作音-大きい」が否定的というように、周囲の語によって評価が変化する例に対応することができた。

● 名詞「起動」を含む素性

「起動」「起動する」を含む素性について、表6に示す。

「起動」という言葉について、その使用法を考察すると、通常ノート PC はシステムが立ち上がる、つまり「起動」するのが当たり前であって、人はあまり当たり前のことには言及しない。したがって、あらためて「起動」という言葉が使われる際には、「起動しなく」なった場合が多いと考えられる。したがって、考察どおり否定的な評価表現であることが多いが、例えば、「起動-

(注10): 特に否定的な評判では精度の向上も見込まれる

(注11): <http://chasen.org/~taku/software/TinySVM/>

早い」のように肯定的な意味を持つ語と伴って現れた場合には肯定的な評価表現となっている。

- 名詞「キーボード」を含む素性

名詞「キーボード」を含む素性について、表7に示す。

ノートPCのキーボードのタッチにこだわる人々も少なくないということは既存の事実である。そこで「キーボード」を含む素性を調べてみると確かに、キーボードの打ちやすさ、打ちづらさがドメイン知識を反映した素性と成っていることが分かる。

5.2 エラー分析に関する考察

抽出エラーとなってしまった例について、分析を行ったところ、エラーとなる原因としては大きく分けて以下の2点が考えられる。

まずは、評判として抽出した文が疑問文であった場合である。

例) そろそろ買っておくのが良い選択なのでしょう
か???

Pentiumの方が速いということはないですか?

本稿で提案した手法では、語の出現のみを考えており話者がどのような意図で文を記しているかは考慮していなかった。疑問文や伝聞・推定の意図で書き込まれた文の形式はドメインに依存せず、共通の構造を持つと考えられる。したがって、ヒューリスティクス等を構築し書き手の意図を考慮する必要がある。

もうひとつは、エラーの割合として前者よりは少ないが、形態素解析等の処理の前提となるツール自体のエラーである。この原因はさらに2つに分けることが出来る。

ひとつは、茶筌の形態素解析自体が誤っている場合である。ただし、この場合は、常に同じ解析誤りを行っている限りは訓練の際も実際の抽出の際も、同様に扱われるので、精度への影響は小さくなる。もうひとつは、茶筌への入力文自体が誤っている、もしくは当て字や略語を含み解析が困難な場合である。

例) 今のマシン (T93B) が古くだったので買い換えるつもりです。(古くなった?)

ソ - (ソニー?), パナ (パナソニック?) など

この問題については、前者よりも解決が困難である。

6. Webへの適用

本章では、我々が試作しているWebからの評判抽出システムに関する報告および、事例研究について報告する。

6.1 システムの概要

本報告の手法を基に、藤村[2]で紹介した、試作評判抽出システムの改良を行った。現在の簡単な処理の流れを図4で紹介する。

Crawling部分については、今回はGoogleAPI^[注12]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^[注13]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:

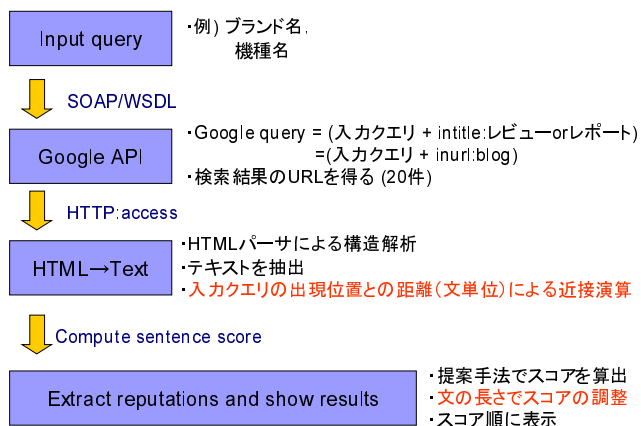


図4 処理の概要



図5 評判抽出結果の例

レビュー OR intitle:レポート」もしくは「inurl:blog」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、クエリとして最近評判の良い、軽量モバイルノートの型番を与えた場合の評判を抽出した際の結果を図5に表示する。

6.2 事例研究

本節では、事例研究として、世間的には「軽くて、バッテリーの持ちが良いモバイルノート」というイメージの製品のWeb上の評判について、ブランド名^[注14]をGoogleにクエリとして用い、検索し得られた計2408ページについて評判抽出を行った。結果として、実際にその製品名が文字列として含まれていたページは1498ページであり、その中から、素性を含んだ文として評判の候補が30587件抽出された。

その文のスコアや含まれている語を中心に分析した結果を図6に示す。スコアの上位の中には、「トラックボールが使いやすいいいマシンでした」というように、現在では採用されていない入力デバイスに対する評価(過去のモデルに対する評価)も

[注12]: <http://www.google.com/apis/>

[注13]: 特に、型番を入れると良い結果が得られやすい

[注14]: ブランド名は英語表記、およびカタカナ表記ともにクエリとして採用している

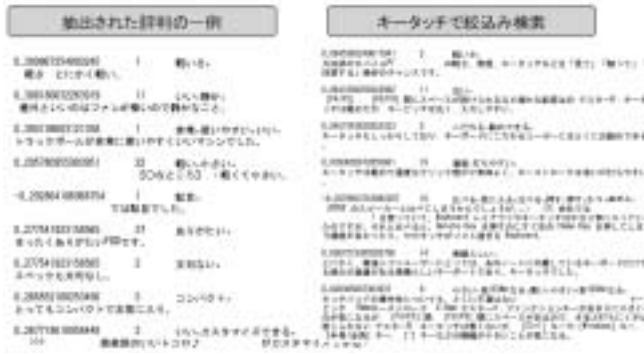


図6 抽出結果

得られている。また、この評判の候補の中からさらにキーワード検索を行うことによって、例えば「キータッチ」で絞り込むことで対象のキーワードに対する評価のみを取り出すことも可能である。

7. おわりに

本稿では、文節 n-gram を素性として用いることで、単語レベルの素性よりも評価表現辞書の手による分析が容易になり、周囲の語によって評価が反転するような語も評価表現として取り込むことができること、および文節 n-gram 素性を用いることで、評判分類・評判抽出における再現率を向上させることができることを示した。

以下、今後の課題について列挙する。

- コーパスの作成支援手法、もしくは少量コーパスでの精度の向上

単語 n-gram の素性を用いることで、単語レベルの素性を作成する場合と比較してより大量のコーパスが必要になると考えられる。表 2 において four-gram のスコアの絶対値は unigram や bigram と比較して小さいこともコーパスの量が不足していることを表していると考えられる。

あらかじめ肯定の評判と否定の評判を用意することによって、ドメイン依存の問題は回避されるが、結果的にコーパス作成が容易ではないという問題が残ってしまう。そこで、今後は少量のコーパスからタグなしテキストのタグを予想しコーパスとして再度利用する手法を用いることで少量のコーパスでも十分な精度が得られるような手法について検討する、もしくは、簡単にコーパスを作成できるような手法を検討する必要がある。

- 類語の処理に関する検討

例) ファンの音がうるさい

ファンの騒音がやかましい

上記の例のようにほぼ同じ意味であるのに、現状の評価表現辞書の構築法では全く違う素性として扱われてしまう。このことがコーパスの量を必要とする原因ともなっている。そこで、既存のソーラスを基に対象としているドメインに対応し類語の処理法について検討する必要があると考えている。

- ドメインに依存しないヒューリスティックスの構築疑問文や書き手の意図が推定・伝聞である文などを判別するための

ルールなどドメインに依存しない、幾らかのルールについてはあらかじめヒューリスティックに構築することで精度の向上が見込まれる。今後はドメインに依存しないヒューリスティックスについては実装が必要であると考えている。

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] 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. 電子情報通信学会技術研究報告, 信学技報 Vol.104 No.177, pp.141-146, 2004
- [3] 乾孝司, 乾健太郎, 松本裕治: 出来事の望ましさを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] Taku Kudo, Yuji Matsumoto. A Boosting Algorithm for Classification of Semi-Structured Text, EMNLP2004
- [6] 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.
- [7] Takeshi Sagara, Masaru Kitsuregawa. Yellow Page driven Methods of Collecting and Scoring Spatial Web Documents. Workshop on Geographic Information Retrieval SIGIR 2004, pp.4-8, 2004.7
- [8] 鈴木泰裕, 高村大也, 奥村学. Weblog を対象とした評価表現抽出. 人工知能学会研究会資料 SIG-SW& ONT-A401-02, 2004.7
- [9] 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.
- [10] 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. 人工知能学会誌, pp.317-323, 2004.5
- [11] 峠泰成, 大橋一輝, 山本和英. 繰り返し学習を用いた話題に順応する意見文抽出. 情報処理学会 研究報告, FI77-5, 2004.11
- [12] 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

Appendix

ここでは、本文中で示したとおり unigram から four-gram に
おいて、信頼性の高かった素性について、肯定・否定各 5 つほ
ど記した表 8~11 を記す。

表 8 スコアの大きい素性 (unigram-1 文節)

(Positive)	頻度 (肯定)	頻度 (否定)	スコア	χ^2 乗値
良い	220	24	0.767	129.7
満足する	135	1	0.982	111.0
いい	210	29	0.714	111.0
静か	114	9	0.826	73.27
快適	65	2	0.928	49.04
(Negative)	頻度 (肯定)	頻度 (否定)	スコア	χ^2 乗値
悪い	3	75	-0.934	81.19
出す	10	76	-0.803	65.32
壊れる	3	53	-0.941	60.80
交換する	7	51	-0.79	43.02
遅い	3	35	-0.867	33.71

表 9 スコアの大きい素性 (bigram-2 文節)

(Positive)	頻度 (肯定)	頻度 (否定)	スコア	χ^2 乗値
音-静か	23	1	0.900	16.47
いい-感じ	16	0	0.929	13.30
買い物-する	15	0	0.925	12.46
使い-やすい	15	0	0.925	12.46
液晶-綺麗	15	0	0.925	12.46
(Negative)	頻度 (肯定)	頻度 (否定)	スコア	χ^2 乗値
修理-出す	2	55	-0.941	60.80
気-なる	38	120	-0.569	57.02
音-する	6	30	-0.715	21.05
悪い-思う	0	10	-0.923	12.03
新品-交換する	0	9	-0.15	10.87

表 10 スコアの大きい素性 (trigram-3 文節)

(Positive)	頻度 (肯定)	頻度 (否定)	スコア	χ^2 乗値
買い物-する-思う	6	0	0.832	4.979
いい-買い物-する	5	0	0.805	4.149
良い-買い物-する	4	0	0.768	3.318
気-なる-ない	3	0	0.713	2.488
満足-いく-買い物	3	0	0.713	2.488
(Negative)	頻度 (肯定)	頻度 (否定)	スコア	χ^2 乗値
音-気-なる	5	22	-0.680	14.14
ファン-回る-頻度	0	4	-0.828	4.827
場合-気-なる	0	4	-0.828	4.827
する-ファン-回る	0	3	-0.783	3.620
ファン-音-うるさい	0	3	-0.783	3.620

表 11 スコアの大きい素性 (four-gram-4 文節)

(Positive)	頻度 (肯定)	頻度 (否定)	スコア	χ^2 乗値
増設する-解決-思う-満足する	2	0	0.623	1.658
ドット-抜ける-ない-いい-感じ	2	0	0.623	1.658
ファン-回る-静か-いい	2	0	0.623	1.658
解決-思う-満足する-使用する	2	0	0.623	1.658
音-気-なる-快適	2	0	0.623	1.658
(Negative)	頻度 (肯定)	頻度 (否定)	スコア	χ^2 乗値
ファン-音-気-なる	3	10	-0.600	5.217
仕様当初-使う-にくい-かもする-ない	0	2	-0.706	2.413
多い-修理-いつ-かかる	0	2	-0.706	2.413
故障箇所-多い-修理-いつ	0	2	-0.706	2.413
普通-使う-壊れる-ない	0	2	-0.706	2.413