

レビュー記事からの評判情報抽出の一手法の提案

梅村 真史

豊橋技術科学大学

ume@smlab.tutkie.tut.ac.jp

酒井 浩之

豊橋技術科学大学

sakai@smlab.tutkie.tut.ac.jp

増山 繁

豊橋技術科学大学

masuyama@tutkie.tut.ac.jp

1 はじめに

ブログや Web 掲示板などには、電化製品等の製品に関する個人の意見が数多く存在している。これらは、消費者が購入する製品を選ぶときや、企業が次期製品開発の計画を練るときなど、様々な場面で貴重な情報源となる。従って、ブログや Web 掲示板などから消費者の製品に対する意見を収集し、よりきめ細かく消費者の意見を分析する必要がある。しかしながら、ユーザの製品に対する意見は、Web 掲示板等の人手では読み切れない莫大な量のテキスト中に存在するために、収集することが困難となる。本研究では、対象を Web 掲示板に限定し、掲示板に存在する製品に関する意見（以降、これを評判情報と呼ぶ）を抽出することを目的とする。

本研究で抽出する評判情報は、次のような製品に対する意見を含む文として定義する。

A 社のカメラの解像度が素晴らしいです。

本研究では、このような評判情報を抽出するにあたり、「評価項目」と「評価表現」と呼ぶ文の一部に着目する。「評価項目」とは、多くの評判情報に含まれるユーザが評価の対象としている項目を表し、上記の評判情報の例では「解像度」がこれにあたる。また、「評価表現」とは、多くの評判情報に含まれる、評価項目への意見・感想を表し、上記の評判情報の例では「が素晴らしいです」がこれにあたる。本研究では、評価項目と評価表現を自動的に同定し、この 2 つの要素を含む文を評判情報として抽出する。

以下、本稿では、関連研究として、製品等の評判を取り扱う研究や本研究と同様に自動的に表現や意味関係を抽出する手法について述べる。その後、我々が提案する評判情報の抽出手法について説明し、実際に kakaku.com^{*1}を対象として、既存手法を本研究に適用させ比較実験を行う。

2 関連研究

先行研究では、辞書やパターンを用いて評判情報を抽出する手法が提案されている。立石ら [7] は、製品カテゴリごとに評価表現の辞書を人手により作成し、入力された商品名と評価表現を近接演算することで製品に対する意見を抽出する手法を提案している。さらに、立石らは構文的な特徴から意見らしさを定義し、抽出した意見のフィルタリングを行っている。立石らの研究は、評価表現辞書を人手で作成する必要があり、多大な時間や労力を要する。また、新たな製品カテゴリが生まれたり、製品カテゴ

リの分類が変化したりした場合、その都度、評価表現辞書を作らなければならない。小林ら [2] は、意見の主要部分は <対象, 属性, 評価> の 3 要素からなると考え、それぞれの要素の辞書を構築する手法を提案している。各要素の抽出は、共起パターンを人手によって与えることで行っている。小林らの研究も人手を用いているので、立石らの研究と同様の問題が考えられる。Hu ら [1] は、抽出対象の素性を名詞と名詞句、形容詞に限定し、相関ルールマイニングと出現頻度により、辞書を構築する手法を提案している。しかしながら、素性の獲得方法は人手で定めた出現頻度のみを用いており、適切な素性が不適切な素性かの選別を行っていない。また、抽出する素性の品詞を限定しているため、名詞や名詞句、形容詞以外の素性を抽出することができない。

一方、抽出対象は評判情報ではないが、自動的に表現や意味関係を抽出する手法も提案されている。Sakai ら [6] は、企業の業績発表記事から業績要因を表す表現を共通頻出表現と手がかり表現の 2 つに分割し、少数の手がかり表現からエントロピーに基づいて双方を自動的に抽出している。Pantel ら [5] は、少量のシードとインスタンスから相互情報量に基づいて is-a や part-of などの意味関係を自動的に獲得している。また、Komachi ら [3] は Pantel らの手法をグラフ解析として定式化し、グラフ解析手法におけるトピックドリフト現象と、ブートストラップにおける意味ドリフトとの関連性について述べている。

本研究では、抽出対象となる評判情報に含まれる評価項目と評価表現に対して、エントロピーに基づいて重み付けを行い、重み付けされた評価項目と評価表現から、新たな評価項目と評価表現を自動的に獲得する。そして、獲得した評価項目と評価表現を用いて、評判情報を自動的に抽出する手法を提案する。また、Sakai らの手法と Pantel らの手法を本研究に適用させ、本手法との比較を行う。

3 評判情報抽出手法の概要

手法の説明にあたり、評価項目と評価表現を以下のように定義する。

評価項目: 製品やサービスなどの様々な項目を表している文節の最後尾から、助詞や形式名詞「の」を削除した形態素列（詳細は後述）

評価表現: 評価項目に係る文節に、その評価項目を構成する最後尾の文節に含まれる助詞や形式名詞「の」を追加したポジティブやネガティブ、要望を表す形態素列（詳細は後述）

評価項目と評価表現の例を以下に示す。枠で囲まれている表現が評価項目であり、下線が引いてある表現が評価表現になる。

*1 <http://www.kakaku.com>

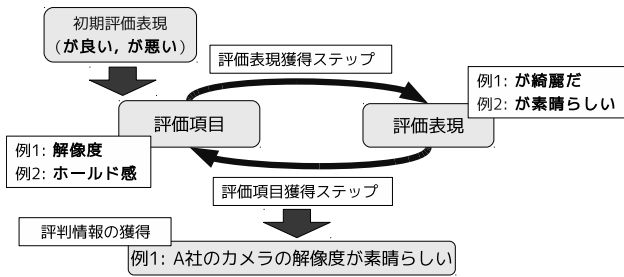


図1 評価表現と評価項目の自動獲得

- A 社のカメラの **解像度** が良いです。
 B 社の液晶の **解像度** は素晴らしいです。
 C 社のレンズの **解像度** は悪いです。

「A 社のカメラの解像度が良いです。」という文において、これを文節に区切ると「A 社の」「カメラの」「解像度が」「良いです。」となる。この文に含まれている評価項目となる文節は「解像度が」となり、この文節から助詞「が」を削除した「解像度」が評価項目となる。また、「良いです。」から句点を削除した「良いです」に削除された助詞「が」を追加した「が良いです」が評価表現となる。評価項目を構成する文節から、助詞や形式名詞「の」を削除した理由は、同じ評価項目を含む文節でありながら、評価表現が異なるため評価項目が変化することを防ぐためである。例えば、「A 社のカメラの解像度が良いです。」の場合、評価項目を構成する最後尾の文節は「解像度が」となるが、そこから助詞「が」を削除することで評価項目は「解像度」となり、評価表現は「が良いです」となる。それに対し、「B 社のカメラの解像度は素晴らしい」の場合、評価項目を構成する最後尾の文節は「解像度は」となるが、そこから助詞「は」を削除することで評価項目は「解像度」となり、評価表現は「は素晴らしいです」となる。このように、同じ評価項目「解像度」を含む文でありながら、評価表現が異なるため評価項目を構成する最後尾の文節が異なるが、助詞や形式名詞「の」を削除することで評価表現が異なる場合でも同じ評価項目を獲得することが可能となる。また、評価項目の取得の際に削除された助詞や形式名詞「の」を評価項目が係っている文節に追加することで、評価項目が前に出現する可能性が高い評価表現となる。

本手法の概要を以下に示す (図1 参照)。

- Step 1 1 つ以上の評価表現を最初に人手で与え、それに係る*2評価項目を獲得する。
 Step 2 獲得した評価項目から、新たな評価表現を獲得する。
 Step 3 獲得した評価表現から、新たな評価項目を獲得する。
 Step 4 Step 2, 3 を予め定められた回数だけ繰り返す。
 Step 5 Step 2, 3 で求められた評価項目候補と評価表現候補の再計算を行う (詳細は後述)。
 Step 6 評価項目と評価表現を用いて、評判情報を抽出する。

次節で、評価表現から評価項目を獲得する手法について述べる。

3.1 評価項目の獲得と選別

評価表現から新たな評価項目を獲得するための手法について述べる。まず、評価表現を含んでいる文を抽出し、その中で評価表現に係っている文節を獲得する。そして、その文節に対して、係っている評価項目を構成する文節に含まれる助詞や形式名詞「の」を削除し、それを評価項目候補とする。しかしながら、獲得した評価項目候補には不適切な評価項目候補も含まれる。そこで、獲得された評価項目候補の中から適切な評価項目候補を選別し、それを評価項目として獲得する。具体的には、適切な評価表現は適切な評価項目に係られるという仮定に基づき、評価項目の重みを求める。評価項目の重みは式1で求める。

$$W(y) = \sum_{x \in X} W(x) \times \frac{f(y, x)}{N_t(x)} \times H(y) \quad (1)$$

ただし、

- $W(x)$: 評価表現 x に与えられている重み
 $f(y, x)$: 評価表現 x に係る評価項目 y の出現回数
 $N_t(x)$: 評価表現 x に係る評価項目の総数
 $H(y)$: 評価項目 y の重要度
 X : 評価項目 y が係る評価表現の集合

ここで $H(y)$ は、様々な評価表現に係っている評価項目は重要であるという仮定のもと、評価項目が評価表現に係る確率に基づくエントロピーとなる。これは、評価項目の重要度をあらかじめ求めておき、評価項目の重要度に基づいて評価項目の重みを求めることで、適切な評価項目により高い重みが付くことが期待できる。評価項目が評価表現に係る確率に基づくエントロピーは式2で求める。

$$H(y) = -\frac{1}{\log_2(N_d(y) + 1)} \sum_{x \in X} P(y, x) \log_2 P(y, x) + 1 \quad (2)$$

$$P(y, x) = \frac{f(y, x)}{N_t(y)} \quad (3)$$

ただし、

- $P(y, x)$: 評価項目 y が評価表現 x に係る確率
 $N_t(y)$: 評価表現に係る評価項目 y の総数
 $N_d(y)$: 評価項目 y が係る評価表現の異なり数

また、最初に人手で与えられる評価表現の重みは、 $W(x) = 1/X$ とし、このとき X は人手で与えられた評価表現の異なり数となる。

最後に、重みが求められた評価項目候補を、重みの大きい順にソートし、上位 N 個の評価項目候補を評価項目として獲得する。

3.2 評価表現の獲得と選別

評価項目から新たな評価表現を獲得するための手法について述べる。まず、評価項目を含んでいる文を抽出し、その中で評価項目に係っている文節を獲得する。そして、その文節に対して、係っている評価項目の文節に含まれる助詞や形式名詞「の」を追加し、それを評価表現候補とする。しかしながら、獲得した評価表現候補には不適切な評価表現候補も含まれる。そこで、獲得された評価表現候補の中から適切な評価表現候補を選別し、そ

*2 係り受け解析器“Cabocha”を使用
<http://chasen.org/taku/software/cabocha/>

れを評価表現として獲得する。具体的には、適切な評価項目は適切な評価表現に係るという仮定に基づき、評価表現の重みを求める。評価表現の重みは式 4 で求める。

$$W(x) = \sum_{y \in Y} W(y) \times \frac{f(x, y)}{N_t(y)} \times H(x) \quad (4)$$

ただし、

$W(y)$: 評価項目 y に与えられている重み
 $f(x, y)$: 評価項目 y によって係られる評価表現 x の出現回数
 $N_t(y)$: 評価項目 y によって係られる評価表現の総数
 $H(x)$: 評価表現 x の重要度
 Y : 評価表現 x に係る評価項目の集合

ここで $H(x)$ は、様々な評価項目に係られている評価表現は重要であるという仮定のもと、評価表現が評価項目に係られる確率に基づくエントロピーとなる。これは、評価表現の重要度をあらかじめ求めておき、評価表現の重要度に基づいて評価表現の重みを求めることで、適切な評価表現により高い重みが付くことが期待できる。評価表現が評価項目に係られる確率に基づくエントロピーは式 5 で求める。

$$H(x) = -\frac{1}{\log_2(N_d(x) + 1)} \sum_{y \in Y} P(x, y) \log_2 P(x, y) + 1 \quad (5)$$

$$P(x, y) = \frac{f(x, y)}{N_t(x)} \quad (6)$$

ただし、

$P(x, y)$: 評価表現 x が評価項目 y によって係られる確率
 $N_t(x)$: 評価項目に係られる評価表現 x の総数
 $N_d(x)$: 評価表現 x に係る評価項目の異なり数

最後に、重みが求められた評価表現候補を、重みの大きい順にソートし、上位 N 個の評価表現候補を評価表現として獲得する。

3.3 評価項目候補の重みの再計算

反復して評価項目候補の重みを求めた後に、各反復で求められた評価項目候補の重みの再計算を行う。具体的には、各反復の評価項目の獲得処理において出現する評価項目候補は評価項目である可能性が高いと考え、各反復の評価項目の獲得処理で計算された評価項目候補の重みの和を求める。例えば、反復の 1 回目で「が良い」「が悪い」といった評価表現から評価項目候補「画質」が出現し、反復の 2 回目で「が綺麗です」「がイマイチ」といった評価表現から出現する評価項目候補が「画質」であるなら、反復の 1 回目と 2 回目の「画質」の重みの和は高い重みを持つ。各反復の評価項目の獲得処理で出現する評価項目候補の重み付きの和は式 7 で求める。

$$W_t(x) = \sum_{i=1}^n W(x_i) \quad (7)$$

ここで、 n は反復回数をあらわし、 $W(x_i)$ は各反復の評価項目の獲得処理で求められた評価表現候補の重みとなる。

3.4 評価表現候補の重みの再計算

反復して評価表現候補の重みを求めた後に、各反復で求められた評価表現候補の重みの再計算を行う。具体的には、評価項目候補の重みの再計算と同様に、各反復の評価表現の獲得処理において出現する評価表現候補は評価表現である可能性が高いと考え、各反復の評価表現の獲得処理で計算された評価表現候補の重みの和を求める。

$$W_t(y) = \sum_{i=1}^n W(y_i) \quad (8)$$

ここで、 n は反復回数をあらわし、 $W(y_i)$ は各反復の評価表現の獲得処理で求められた評価表現の重みとなる。

3.5 評判情報の獲得

反復して重みが計算された評価項目と評価表現を、重みの大きい順にソートし、重みの大きい順に K 個の評価項目と評価表現を評判情報の獲得に使用する。そして、 K 個の評価項目と評価表現を用いて、評価項目が評価表現に係っている表現を含む文を評判情報として抽出する。

例えば、評価項目として「画質」が獲得され、評価表現として「が綺麗だ」が獲得された場合、「A 社の液晶テレビの画質が綺麗だ。」という文は「画質」が「が綺麗だ」に係っているため、この文は評判情報として獲得される。

4 評価実験

本研究では、kakaku.com のカテゴリ、「デジタルカメラ」のレビューを用いた。各文書の html タグを除去して文単位に分割し、係り受け解析を行った文書、124,484 文を使用した。また、Pantel ら [5][4] の手法と Sakai ら [6] の手法を本研究に適用させ、提案手法との比較を行った。

提案手法と Pantel らの手法、Sakai らの手法の初期評価表現は「が良い」「が悪い」とし、提案手法と Pantel らの手法の反復回数は 10 回、Sakai らの手法の反復回数は 5 回とした。評価には、コーパス中からランダムに抽出した 2000 文を評価用データセット（正例 749 文、負例 1251 文）として使用する。ここで、正例は評判情報を含む文、負例は評判情報を含まない文とする。また、評価尺度には精度と再現率、F 値（調和平均）を用いた。

提案手法と Pantel らの手法について、評価項目と評価表現の獲得個数、上位 N を 50 から 100 まで 10 ずつ獲得個数を変化させて実験を行った。また、評判情報の獲得に使用する評価項目と評価表現の数 K を 200 から 500 まで 100 ずつ変化させたときの結果を表 1, 2, 3, 4 に示す。Sakai らの手法については、閾値を 0.4 から 0.6 まで 0.1 ずつ変化させて実験を行った結果を表 5 に示す。

表 1 から表 4 の結果より、精度がもっとも高かったのは提案手法（再計算あり）の $K = 200$, $N = 50$ の 0.794 で、再現率がもっとも高かったのは提案手法（再計算なし）の $K = 500$, $N = 100$ の 0.300 となった。また、表 5 より精度がもっとも高かったのは閾値が 0.6 の 0.867 で、再現率がもっとも高かったのは閾値が 0.4 の 0.258 となった。また、実験結果より、F 値がもっとも高かったのは提案手法（再計算なし）の $K = 500$, $N = 100$ の 0.418 となった。

表1 $K = 200$

上位 N	提案手法 (再計算なし)			提案手法 (再計算あり)			Pantel らの手法		
	精度	再現率	F 値	精度	再現率	F 値	精度	再現率	F 値
50	0.729	0.166	0.270	0.794	0.150	0.252	0.757	0.142	0.238
60	0.719	0.188	0.298	0.753	0.155	0.257	0.769	0.151	0.252
70	0.713	0.184	0.292	0.741	0.160	0.263	0.762	0.150	0.250
80	0.721	0.190	0.300	0.735	0.163	0.267	0.763	0.155	0.257
90	0.712	0.188	0.298	0.720	0.175	0.281	0.773	0.159	0.264
100	0.715	0.191	0.301	0.710	0.174	0.279	0.761	0.158	0.261

表3 $K = 400$

上位 N	提案手法 (再計算なし)			提案手法 (再計算あり)			Pantel らの手法		
	精度	再現率	F 値	精度	再現率	F 値	精度	再現率	F 値
50	0.695	0.228	0.344	0.703	0.222	0.337	0.700	0.206	0.318
60	0.696	0.256	0.375	0.703	0.240	0.358	0.686	0.239	0.354
70	0.704	0.267	0.387	0.698	0.247	0.364	0.674	0.232	0.346
80	0.694	0.267	0.386	0.690	0.247	0.363	0.685	0.247	0.363
90	0.690	0.271	0.389	0.684	0.248	0.364	0.678	0.246	0.361
100	0.689	0.272	0.390	0.686	0.242	0.357	0.675	0.250	0.365

表5 Sakai らの手法の結果

閾値	精度	再現率	F 値
0.6	0.867	0.087	0.154
0.5	0.757	0.150	0.250
0.4	0.594	0.258	0.360

5 考察

提案手法と Pantel らの手法, Sakai らの手法の比較を行った結果, 有意な差はみられなかった. しかしながら, 評判情報の抽出に使用する評価項目と評価表現の数を増やしたとき, 表3と表4, 表5より, 本手法が Pantel らの手法や Sakai らの手法より精度, 再現率, F 値がいずれもわずかに高かった.

精度に有意な差が見られなかった理由として, 不適切な評価項目や評価表現が獲得されたため, 不適切な評判情報が抽出されたことが原因である. 例えば, 評価項目として「気持ち」や評価表現として「を求めて」が取得されたため「何枚撮っても気持ちがいい。」や「最適なバランスを求めて, 色々試していきたいと思っています。」といった文が抽出され, これが正解率を下けている要因の一つと考える. また, 再現率に有意な差が見られなかった理由として, 文書中に出現する頻度が少ないため, 獲得されない評価項目や評価表現が多いことが原因である. 例えば, 「ポートレート」や「は遅れ気味」といった評価項目や評価表現は, それぞれ1回しか出現してない. 特に評価表現は, 書き手によって異なるため, 同様の評価表現が出現しにくくなり, 出現頻度が低くなると考えられる. これは, 新聞記事などとは異なり, レビュー記事には様々な書き手が存在し, 書き方が統一されていないことが原因であると考えられる. そのため, 低頻度の評価項目や評価表現を獲得するのは難しいと考えられる.

6 まとめと今後の課題

本研究では, レビューから「液晶テレビの映りがきれいだ」「A社のカメラの解像度が良い」といった製品に関する評判情報を自動的に獲得した. 抽出対象の表現を「評価項目」と「評価表現」に分割し, この二つの要素をコーパスから自動的に獲得することで, 多数の評判情報の抽出を行う. 例えば, 「液晶テレビの映り」や「カメラの解像度」が評価項目となり, 「がきれいだ」

表2 $K = 300$

上位 N	提案手法 (再計算なし)			提案手法 (再計算あり)			Pantel らの手法		
	精度	再現率	F 値	精度	再現率	F 値	精度	再現率	F 値
50	0.701	0.207	0.320	0.714	0.187	0.296	0.735	0.186	0.296
60	0.700	0.224	0.340	0.701	0.200	0.312	0.726	0.195	0.307
70	0.694	0.234	0.350	0.705	0.207	0.320	0.707	0.194	0.304
80	0.699	0.230	0.346	0.697	0.206	0.318	0.700	0.194	0.303
90	0.696	0.230	0.345	0.694	0.206	0.317	0.703	0.199	0.310
100	0.688	0.227	0.341	0.699	0.211	0.323	0.700	0.212	0.326

表4 $K = 500$

上位 N	提案手法 (再計算なし)			提案手法 (再計算あり)			Pantel らの手法		
	精度	再現率	F 値	精度	再現率	F 値	精度	再現率	F 値
50	0.698	0.259	0.378	0.711	0.240	0.359	0.692	0.243	0.360
60	0.683	0.284	0.402	0.708	0.259	0.379	0.675	0.260	0.376
70	0.687	0.296	0.414	0.696	0.268	0.387	0.677	0.260	0.376
80	0.681	0.296	0.413	0.688	0.271	0.389	0.661	0.279	0.392
90	0.679	0.296	0.413	0.690	0.274	0.392	0.672	0.276	0.392
100	0.686	0.300	0.418	0.687	0.275	0.393	0.675	0.286	0.402

や「が良い」が評価表現となる. そして, 最初に少数の評価表現「が良い」「が悪い」を手で与え, コーパスの統計情報を使用して反復して評価項目と評価表現のペアを取得することで, 自動的に評判情報の抽出が可能である.

kakaku.com の「デジタルカメラ」のレビューに本手法と Pantel らの手法, Sakai らの手法を適用した結果, 有意な差は見られなかった. しかしながら, 評判情報の抽出に使用する評価項目や評価表現の数を増やしたとき, 本手法が Pantel らの手法や Sakai らの手法より精度, 再現率, F 値がいずれもわずかに高かった.

本研究の今後の課題として, 不適切な評価項目や評価表現の除去と文書中に出現する頻度が少ない評価項目と評価表現の抽出が考えられる.

参考文献

- [1] Minqing Hu and Bing Liu. Mining opinion features in customer reviews. In *American Association for Artificial Intelligence*, pp. 755–760, 2004.
- [2] 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. 情報処理学会研究報告, pp. 77–84, 2003.
- [3] Mamoru Komachi, Taku Kudo, Masashi Shimbo, and Yuji Matsumoto. Graph-based analysis of semantic drift in espresso-like bootstrapping algorithms. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, pp. 1011–1020, 2008.
- [4] Patrick Pantel and Deepak Ravichandran. Automatically labeling semantic classes. In *Proceedings of HLT/NAACL-04*, pp. 231–328, 2004.
- [5] Patrick Pantel and Marco Pennacchiotti. Espresso: Leveraging generic patterns for automatically harvesting semantic relations. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the ACL*, pp. 113–120, 2006.
- [6] Hiroyuki Sakai and Shigeru Masuyama. Cause information extraction from financial articles concerning business performance. *IEICE Trans. Information and Systems*, Vol. E91-D, No. 4, pp. 959–968, 2008.
- [7] 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, pp. 75–82, 2001.