

# 評判情報の注釈付けタグセットの提案

宮崎林太郎

横浜国立大学 大学院  
環境情報学府

前田直人

横浜国立大学 工学部  
電子情報工学科

森辰則

横浜国立大学 大学院  
環境情報研究院

E-mail:{rintaro, n-maeda11, mori}@forest.eis.ynu.ac.jp

## 1. はじめに

近年, Weblog 等の普及により, 個人が製品やサービスなどに対するレビューを公開する機会が増えている. また, 価格.com<sup>1</sup>やAmazon<sup>2</sup>のような, 製品レビューを集約して扱うサイトも登場して久しい. これに伴い, 文中から筆者の主観的な記述を抽出・解析する試みが行われている.

本研究の目的は, 評判情報抽出タスクに関する研究を推進するにあたって必要不可欠と考えられる, 評判情報コーパスを作成することにある. 我々はそのために, 評判情報コーパスの作成を目的とした評判情報モデルの策定と, タグセットの仕様を検討している. 本論文では, 同モデルの提案を行うと共に, コーパスの作成実験を行い, その過程で得られた知見を報告する.

## 2. 関連研究

評判情報の抽出に関する初期の研究に立石ら[1]のものがある. この研究では, 用意された評価表現辞書を用いて, 対象物と評価表現を含む一定の範囲を, 意見として抽出している. その後も, 同様の評価表現辞書を用いて評判情報抽出を試みる研究が行われている. 一方で機械学習手法を用いて評判情報を抽出する研究[3]も行われている.

また, 評判情報に関連する研究として, 評価の肯定/否定の極性分類が挙げられる. 極性分類に対しても, 評価表現辞書を用いる手法と, 機械学習手法を用いる手法[5][6]が研究されている.

これら, 機械学習手法における訓練データ, また他の評判情報抽出タスクにおいても, 評価データなどに用いるコーパスの必要性が高まっている.

評判情報のモデル化においては, 村野ら[2]が挙げられる. この研究では, 主観的評価の構成要素を細かく分類している. 構成要素を[対象][比較対象][評価][項目][様態]とし, それぞれの辞書を用いて抽出を行っている.

また最近では小林ら[4]が, 意見情報の構成要素を[態度][評価][根拠][理由関係]とする研究を行っている. この研究では[評価]についてはさらに細かく, [評価主体][評価対象][評価視点][比較対象][評価条件]の構成要素からなる考え, コーパスの作成を行っている.

上記の先行研究に対して, 本論文の貢献は, 1) 対象となる製品やサービスの様態と, 話者の評価を2層に分離することにより評判情報を精密に記述するとともに, 2) 従来の評判情報モデルでは取り上げられなかった, 文中に陽に現れないオントロジーに関する情報を考慮することができるモデルを提案する点にある.

## 3. 評判情報のモデル化とタグセットの提案

### 3.1 評判情報のモデル化

評判情報とは, 製品やサービスに対する個人の主観的なことが述べられたものである. 本論文では評判情報の基本構成要素は次の各項の組であると定義する.

[項目]: 製品やサービスを構成する要素を意味する概念クラスやそのインスタンスを指示する表層表現.

[属性]: 項目の様態を表す観点.

[属性値]: 属性に対する様態の内容.

[評価]: [項目]に対する主観的見解.

[項目]が属する概念クラスは部分-全体関係や上位-下位関係といった階層構造を有する. この情報はいわゆるオントロジー(の一部)を構成するものであるため, 必ずしも文中に陽に現れるものではない. そこで, 我々はオントロジーに関する情報を文中に付与するタグとは別に記述する手法を考える. また, [評価]の理由となるのが[属性]-[属性値]の組である. [属性値]は具体的な数値の場合や, 主観的な表現の場合がある. このモデルで捉えた例文を図1に示す. なお, 各構成要素は文中で省略される場合がある.

{この製品:項目}は{お値段:属性値}が{安く:属性値}で{感激:評価}です.

図1. 評判情報の基本構成要素

本研究で用いる評判情報のモデルを図2に示す. このモデルでは評価の属する層と項目様態を表す層の2層構造にすることで, 対象となる[項目]の様態と, 話者の[評価]を分離することが可能である.

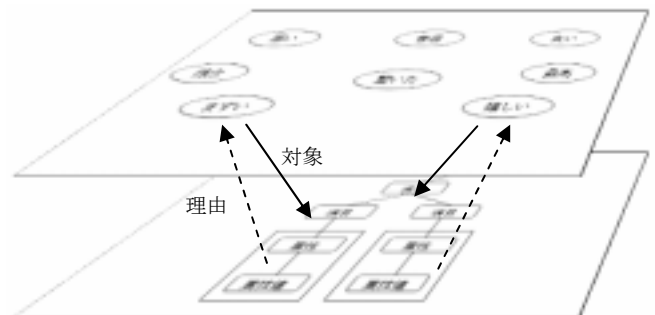


図2. 評判情報の提案モデル

さらに, 基本構成要素の組に加えて, 周辺情報を以下の通り定義する. 周辺情報を付与する対象は文の持つモダリティ情報[7]から定義した. これは, 特に属性値や評価をよ

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

<sup>2</sup> <http://www.amazon.co.jp>

り正確に捉えるために必要である。

[条件]: [属性値]や[評価]がある条件下で述べられている場合. 付与する対象は, 文中で条件を述べている部分全体.  
[比較対象]: [条件]の一部とも考えられるが, 特に重要な情報であると考え, 別個のものとして扱う. 付与する対象は, 文中において, 比較対象となる[項目]や[属性]や[属性値]の部分, またその組が記述されている部分全体.  
[推定]: [属性値]や[評価]を話者が断定保留している場合. 付与する対象は, 文中の[推定]を表す記述部分.  
[伝聞]: 話者が他者からの伝聞で語っている場合. 付与する対象は, 文中の[伝聞]を表す記述部分.  
[否定]: [属性値][評価]の意味・極性を反転している場合. 付与する対象は, 文中における[属性値][評価]に対する否定表現部分.  
[要望]: 感嘆文のような形で要望を述べている場合. 付与する対象は, 要望を述べている文全体.

### 3. 2 提案するタグセット

前節で述べたモデルに基づき作成したタグセットを図3に示す. 前節で述べた構成要素に対応したタグの要素名と属性を示してある.

なお, 各タグにはそれぞれ一意に決まる識別子を id 属性に付与する. これに加えて, 個々の意見文書の単位で話者の記述が「肯定・中立・否定・意見性なし」のどれに該当するかを付与している.

<pre>&lt;item&gt;: 項目   class: この項目が属する概念番号 &lt;attribute&gt;: 属性   pair: 属性-属性値の組番号   target: 属性-属性値を持つ概念番号を参照 &lt;value&gt;: 属性値   pair: 属性-属性値の組番号   target: 属性-属性値を持つ概念番号を参照 condition(任意): 条件がある場合にはその id compare(任意): 比較対象がある場合にはその id &lt;evaluation&gt;: 評価   target: 評価の対象となる概念番号を参照   reason: 評価理由となる属性-属性値の組番号を参照 orientation: 評価が肯定・中立・否定のどれか condition(任意): 条件がある場合にはその id compare(任意): 比較対象がある場合にはその id &lt;condition&gt;: 条件 &lt;compare&gt;: 比較対象 &lt;estimation&gt;: 推定   target: 推定の対象となる属性値・評価を参照 &lt;report&gt;: 伝聞   target: 伝聞の対象となる属性値・評価を参照 &lt;not&gt;: 否定   target: 否定の対象となる属性値・評価を参照 &lt;demand&gt;: 要望</pre>
--

図3. 注釈付けに用いたタグセット

## 4. コーパスの作成実験

今回提案した評判情報モデルとタグセットの問題点を明らかにするために, コーパスの作成実験を行った.

## 4. 1 実験内容

同一の文書群に対して4人の注釈者がタグの付与を行った. また, 同時にオントロジーに関する情報も記述してもらった. 今回使用した文書群は4つの異なるドメインに属する製品に対するレビュー文書である. 本論文ではAmazonのユーザーレビューを用いた. また注釈を行った被験者は情報工学を専攻する学生4名であり, 2名(注釈者1, 注釈者2)は本研究と直接の関連を持つが, 2名(注釈者3, 注釈者4)は直接の関係を持たない.

本論文で用いたトピックを表1に, 注釈付けされた文の実例を図4に示す.

表1. 使用文書のトピック

Topic	Sentence 数
プリンタ	49 文
圧力鍋	49 文
音楽 CD	51 文
ゲームソフト	51 文

```
<sentence id="s06">
最新だからか、<attribute id="a441" pair="p441"
target="c02">印刷スピード</attribute>が<value
id="v451" pair="p441" target="c02">すごく速い
</value>です。
</sentence>
```

図4. 注釈付けされた文の実例

また, 記述されたオントロジー情報の実例を図5に示す. オントロジー情報は[概念クラス]<表層表現>で記述し, “+”を用いて階層が判別できるように記述してもらった.

```
[c01]<非電化製品>
+[c141]<圧力鍋>
++[c02]<この圧力鍋><T-FAL クリップソーシック>
+++[c931]<付録>
++++[c932]<レシピ本><レシピブック>
+++[c1201]<蓋>
```

図5. オントロジー情報の実例

## 4. 2 注釈者間のタグ付与の一致率

今回の実験で各注釈者が付与したタグの総数を表2に示す. 注釈者により同一箇所タグを付与するか否かの判断が異なる場合があることがわかる.

表2. 注釈者ごとのタグ付与の数

	Topic1	Topic2	Topic3	Topic4
注釈者 1	137	122	191	153
注釈者 2	136	94	124	122
注釈者 3	112	59	179	86
注釈者 4	142	138	137	111

注釈者によるタグの有無に関しては, 認識の違いに加え注釈者がどこまで注意深く注釈付けを行うかに拠るところもあるため, 一定の差は生じてしまう. これについては, 全体の文書量を増やすなどしてコーパスとしての有用性を高めることができると思われる.

次にこの中から, 基本構成要素である4つのタグに特に

注目して分析を行った。4つのタグが付与されている文字列のうち、同一部分に同一タグが付与されている一致率を計算した。結果を表3に示す。各タグについて、注釈者4人の組み合わせ(全6通り)におけるκ値の最大値(上段)と最小値(下段)を示してある。

表3. 各タグ付与作業の一致率(κ値)

	Topic1	Topic2	Topic3	Topic4
Item	0.53	0.58	0.62	0.45
	0.33	0.43	0.19	0.19
attribute	0.55	0.69	0.38	0.56
	0.43	0.28	0.11	0.14
value	0.66	0.40	0.37	0.48
	0.41	0.17	0.11	0.16
evaluation	0.81	0.73	0.42	0.70
	0.51	0.29	0.22	0.37

κ値は主観が入る判定が偶然に拠らず一致する割合であり、0.41から0.6の間ならば中程度の一致、0.8を超えるとほぼ完璧な一致と考えられる。

結果を見ると、Topic1・2に比べてTopic3・4におけるκ値が低い。このような結果となる理由としては、対象となる製品が属するドメインの違いが考えられる。実体を持つ製品に比べ、ソフトウェアのような製品のレビューは、概念的な表現や主観的な表現による記述が多くなると考えられる。そのために、Topic3・4における注釈者間の揺れが大きくなっていると考えられる。

#### 4.3 タグの種類による判断の違い

表2からもわかるように、注釈者によって、タグを付与するかしないかの判断は分かれる。しかし、タグが付与されていないというのは、情報が欠落することであり、誤った情報が付与されることとは異なる。そこで、文中の同一部分に対して注釈者によって異なるタグが付与された場合を分析した。結果を表4に示す。

表4. 注釈者により異なるタグが付与された数

付与が異なったタグ	個数
<value>と<evaluation>	55
<item>と<attribute>	15
<condition>と<attribute>	9
<attribute>と<evaluation>	6

同一部分に対して<value>と<evaluate>というタグが付与されている異なりが最も多かった。例文を図6に示す。

付録?のレシピ本は使いやすいです。

例文では「使いやすい」に対して注釈者により<value>と<evaluation>の付与がわかれた。<value>タグを付与した注釈者は、{レシピ本:項目}について{使いやすい:属性}が省略されていて、{使いやすい:属性値}が記述されていると考えた。一方で<evaluation>タグを付与した注釈者は{レシピ本:項目}に対して{使いやすい:評価}が記述されていると考えた。図7の例文もこの問題に関連する。

この製品は大きいので気に入っている。

図7. <value>と<evaluation>に対する分析

例文を本論文のモデルで捉えると{この製品:項目}{大きい:属性}{大きい:属性値}{気に入っている:評価}となる。ここで、「大きい」は主観的な記述とも言えるが、「大きい」はその根拠が話者の外にある場合が多く、「気に入っている」に比べると客観性が高い。「気に入っている」は完全に話者の内部にある主観である。この違いを考慮することにより、異なりを削減することができると考えられる。次に異なるタグが付与されていたものとして<item>と<attribute>がある。この認識の異なりは<value>と<evaluation>の異なりと関係がある。例文を図8に示す。

この商品は、とにかく印刷がきれいです。

図8. <item>と<attribute>の付与の判断が分かれた文

<item>と<attribute>の異なりに注目して、例文に対する注釈付けを考えると{この商品:項目}{印刷:属性}と認識する場合と、{この商品:項目}{印刷:項目}と認識する異なりが生じた。

前述の場合、注釈者は「この商品」が持つ属性として「印刷」があると認識し、後述の場合は「この商品」の機能として「印刷」があると認識している。後述の場合には「この商品」と「印刷」の間の関係は概念クラスの階層関係として外部に記述される。

加えて、<value>と<evaluation>の異なりが発生すると、図9に示す4通りの認識が生じてしまう可能性がある。

{この商品:項目}{印刷:属性}{きれいな:属性値}  
 {この商品:項目}{印刷:属性}{きれいな:評価}  
 {この商品:項目}{印刷:項目}{きれいな:属性値}  
 {この商品:項目}{印刷:項目}{きれいな:評価}

図9. タグの不一致が起こりうる可能性

<item>と<attribute>の異なりについては<value>と<evaluation>の認識の異なりを解消することで、<value>に対する<attribute>の決定を助けることが出来る。

なお、今回の実験では4人の注釈者間において、2つのタグ間の認識の異なりはあっても、3つのタグ以上での異なりは無かった。

#### 4.4 要素間の関係の一致率

3.1で述べたように、評判情報の基本構成は[項目][属性][属性値][評価]の4つ組で表される。作成実験で各注釈者が付与した要素間関係の一致を調べるため、各注釈者によって関連付けられた4つ組の一致の度合いを調べた。

4つ組の要素が全て一致したものは、各注釈者間で要素間の関係が完全一致したものと考えられる。また、部分一致のうち、[属性][属性値][評価]の二つまでが不足しているために部分一致となったものは、要素間の関係については注釈者間で同じ認識ができたものとする。

そこで、各注釈者間の一致の度合いを測るために、「完全一致した組+要素が不足しているために部分一致となった組」を一致しているものとし、一方の注釈者のタグ付与を正解とした時の、もう一方の注釈者のタグ付与の再現率と適合率を調べた。結果を表5に示す。

表 5. 抽出された組の一致の度合い

正解とした注釈者	比較対象の注釈者	再現率	適合率
注釈者 1	注釈者 2	0.30	0.37
注釈者 1	注釈者 3	0.12	0.35
注釈者 1	注釈者 4	0.23	0.25
注釈者 2	注釈者 3	0.15	0.35
注釈者 2	注釈者 4	0.33	0.30
注釈者 3	注釈者 4	0.44	0.17

抽出された組は完全一致した重複を除いた総数で 551 組であった。タグ付与が異なったために完全一致しなかった組の中で、4 つ組のうち 1 つの要素に異なるタグが付与されていたものが 40 組、2 つの要素に異なるタグが付与されていたものは 215 組であった。なお、4 人の注釈者全員が完全に一致した組は 10 組、<value>と<evaluation>のタグ付与が異なるために一致しなかった組は、全部で 17 組であった。

#### 4. 5 認識したオントロジーの相違

今回の提案モデルでは<item>は製品等の要素となる概念が表層に現れたものであると考えている。テキストに現れない事柄の認識であるオントロジーについて調べてみた。今回は階層構造には着目せず、4 人の注釈者が認識した概念クラスの総種類数のうち、注釈者全員が同じ概念クラスとして認識した種類数を一致率として調べた。結果を表 6 に示す。

表 6. オントロジーの一致率

	Topic1	Topic2	Topic3	Topic4
一致率	0.22	0.19	0.32	0.25

この結果では、タグの一致と異なり、主観的な要素の多いドメインを対象とした場合の方が、わずかではあるが結果が優れていた。このような結果となった理由を考えると、ハードウェアのような実態のある製品については、<item>と<attribute>間の認識の揺れから概念構造の認識に違いが出てしまっていると考えられる。

#### 4. 6 注釈者ごとの作業時間

最後に、注釈者による注釈付け作業時間の違いについて分析する。結果を表 7 に示す。

表 7. 注釈者ごとの作業時間(分)

	Topic1	Topic2	Topic3	Topic4
注釈者 1	90	74	127	103
注釈者 2	135	99	118	121
注釈者 3	168	68	90	85
注釈者 4	270	187	196	150

注釈付けは Topic1 から順に行ってもらった。注釈作業に対する“慣れ”による作業時間の変化は当然考えられる。

また 4. 2 節でも述べたように、電化製品のようなトピックよりも、音楽のようなトピックの方がより主観的な記述が豊富であると思われるために、トピックごとの注釈付けの難易度の違いが、作業時間の違いに影響することが考えられる。そこで、タグ 1 つあたりの作業時間を見てみる。

結果を表 8 に示す。

この表からみると、各トピックのタグ 1 つの付与に要する時間は同一注釈者ではそれほど大きく変化していない。タグ付与の時間効率に対するトピックの変化の影響は少ないといえる。対象文書がどの程度注釈付け対象となる要素を含んでいるかが作業時間に大きく関係している。

表 8. 注釈者ごとの作業時間(秒/タグ数)

	Topic1	Topic2	Topic3	Topic4
注釈者 1	39.4	36.4	39.9	40.4
注釈者 2	59.6	63.2	57.1	59.5
注釈者 3	90.0	69.2	68.4	59.3
注釈者 4	88.5	81.3	85.8	81.1

#### 5. おわりに

本研究は評判情報コーパス作成に用いるタグセットの提案を行った。また、そのタグセットを用いてコーパスの作成実験を行った。

評判情報のモデル化の精度については、注釈者によるタグの一致は一致率の高かった注釈者間の  $\kappa$  値で 0.4 から 0.8、要素の組の一致精度は再現率で 0.14 から 0.44、適合率で 0.17 から 0.37 であった。また、特に[属性値]と[評価]の間で注釈者によりタグの認識の異なりが多くなることわかった。一方で、ドメインは作業時間の違いに大きな影響を及ぼさないことがわかった。

より、有用性の高いコーパスを作成するためには、注釈者間のタグ付与の一致率の向上が必要である。解決策として、[評価]となるものをより限定することが考えられる。

今後の研究内容としては、各問題点を検討した上で本格的なコーパスの作成を開始し、コーパスを用いた機械学習による評判情報の抽出を試みることを考えている。

#### 参考文献

- [1] 立石健二, 福島俊一, 小林のぞみ: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 NL144-11, pp.75-82 (2001)
- [2] 村野誠治, 佐藤理史:文型パターンを用いた主観的評価文の自動抽出, 言語処理学会第 9 会 年次大会発表論文集, pp67-70 (2003)
- [3] 飯田龍, 小林のぞみ, 乾健太郎, 松本裕治, 立石健二, 福島俊一:意見抽出を目的とした機械学習による属性-評価値対同定極性反転に対応した評価表現モデル, 情報処理学会研究報告 NL165-4, pp21-28 (2005)
- [4] 小林のぞみ, 乾健太郎, 松本裕治:意見情報の抽出/構造化のタスク使用に関する考察, 情報処理学会研究報告 NL171-18, pp111-118 (2006)
- [5] Peter Turney. Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. In Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), pages 417-424. (2002)
- [6] 高村大也, 乾考司, 奥村学:極性反転に対応した評価値モデル, 情報処理学会研究報告 NL168-22, pp141-148 (2005)
- [7] 益岡隆志:モダリティの文法, くろしお出版 (1991)