

複数注釈者による評判情報コーパスの作成と評判情報の抽出

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 大学院
環境情報学府

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, CGM(Consumer Generated Media)と呼ばれるサービスが広く利用されるようになってきている. CGMには利用者が製品やサービス等に対するレビューを多数公開している. また, 製品の情報やレビューを集約する Amazon や価格.com 等のサイトも日常的に利用されるようになってきている. これに伴って, ネット上に存在するレビューから著者の主観的な見解を評判情報として収集分析することが期待されている.

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, 注釈事例を参照しながら複数注釈者の人手により作成することにある. 本稿では, 1 万文の評判情報コーパスを作成した際に得られた知見と共に, 作成したコーパスを用いた評判情報の抽出実験について報告する.

2. 関連研究

2. 1 評判情報の抽出

評判情報の抽出に関する研究には文書分類的なアプローチを行うものと, 評判情報の構成要素を定義し各構成要素を抽出しようとするものがある.

前者の中には文書分類に関する研究として Turney[1]がある. この研究では, 文中語と評価の種表現となる語との共起頻度を用いて文書全体の評価極性を求めている.

後者の中では村野ら[2]が評価文の文型パターンを整理し, その構成要素ごとに辞書を用意して抽出を行っている. また, 小林ら[6]は評判情報の構成要素をブートストラップ的な手法を用いて収集する研究を行っている.

さらに, 評判情報の構成要素に関する研究においては, 各構成要素の組を同定する研究が行われている. 飯田ら[5]は機械学習手法を用いて属性-属性値対を同定している.

上記の機械学習手法における訓練データや, 評判情報抽出手法における評価データ等に用いるために評判情報コーパスの必要性が高まっていると考えられる.

2. 2 コーパス作成と注釈付け支援ツール

主観的な情報の注釈付けとしては徳久ら[11]があげられる. この研究では漫画におけるキャラクターの表情に対して人手で情緒タグを付与している. また, 正解データを複数注釈者のタグを比較して決定している. 鍛冶ら[12]は HTML 文書中の定型文に着目し, 評価文コーパスの自動構築を行っている.

注釈付け支援ツールに関する研究としては, 高橋ら[9]がアノテーションツール“Tagrin”を作成している. これは, ユーザの定義したチャンクに対するラベル付けや関係の付与を目的としたものである. また, 注釈付け作業における事例の提示に関しては洪ら[4]がある. この研究では, 対話コーパス作成の際に注釈者へ過去の事例を提示することで支援を行っている.

上記の先行研究に対して, 本研究の貢献は 1)後述する 4 つ組の要素からなる評判情報モデルの注釈付けを人手で行うことでコーパス作成を行う, 2)注釈付けの際に辞書等を用いず, 過去の注釈事例を用いることで注釈者間の認識を緩やかに共有しコーパスの精度を保つ, 3)作成したコーパスに対して固有表現抽出で用いられる手法を用いて評判情報の抽出を行う点にある.

3. 評判情報の注釈付け

ある程度の規模を持つコーパスを人手で作成する際には複数注釈者による注釈付けが必要不可欠である. しかし, 複数注釈者による注釈付けでは, 注釈揺れが問題となる. 過去の研究[10]から, 注釈事例の提示が注釈揺れの削減に効果を持つことが明らかとなったため, 注釈時に過去の注釈事例を参照しつつ注釈付けを行いコーパス作成を行った.

3. 1 評判情報のモデル化

我々は評判情報の注釈付けタグセットの提案を行い, 構成要素を定義してきた[7]. このタグセットでは, 製品の様態と評価を 2 層構造として分離し, item-attribute-value-evaluation の 4 つ組として評判情報をモデル化している. 本稿もそれに従ったタグセットを用いた. 本稿で用いた評判情報のタグセットを図 1 に示す. 構成要素に対応した XML タグの要素名と属性情報を示してある.

```
<item> 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
  class : この item が属する概念番号
<attribute> item の様態を表す観点
  pair : attribute-value の組番号
  target : attribute-value を持つ概念番号
<value> attribute に対する様態の内容
  pair : attribute-value の組番号
  target : attribute-value を持つ概念番号
<evaluation> item に対する主観的見解
  target : evaluation の対象になる概念番号
  reason : 評価理由となる attribute-value の組番号
  orientation : 評価が肯定・中立・否定のどれか
```

図 1. 注釈付けに用いたタグセット

加えて, 上記の評判情報がレビュー等に現れる場合には, 一部要素が省略されている可能性があるため, 本モデルでは要素の省略を許している.

また, item 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と考えられる. この情報はオントロジー (の一部) を構成し, 文中に必ずしも陽に現れるものではない. 今回の注釈付けではオントロジー情報の作成を第一には考えていないが, 注釈者にはオント

ロジー情報についても説明を行い、注釈付け作業と同時に記録をしてもらった。

3. 2 注釈付けツール

注釈付け作業を簡便にし、注釈者が注釈事例を提示しながら作業を行うために、注釈付けツールを作成した。

注釈時には事例用コーパス中の文から、注釈付け対象となる文との類似度が高い上位 5 文を提示した。

また、ボタンによるタグの付与や、XML タグの属性値情報を一部自動入力化し、注釈者を支援する。

3. 3 注釈者への事前説明

注釈者には上記の評判情報のモデルについての解説をし、過去の注釈事例から「注釈揺れが起きた実例」「判断が難しかった実例」を示しながら事前説明を行った。

今回のコーパス作成では、注釈者に value タグと evaluation タグから着目して注釈付けを行うように指示した。また、その際に value と evaluation に該当する文字列がどちらも存在せずに item や attribute のみが現れた場合には注釈付けを行わなくてもよい旨を説明した。

また、過去の注釈付け実験においては、似通った部分に注釈付けされているが、その範囲が微妙に異なる注釈揺れが存在した。そのために、今回の注釈付け作業ではタグの範囲の細かい揺れを削減するための規則も追加した。追加した規則を図 2 に示す。さらに、モダリティ[13]に関連する部分も過去の注釈付けで注釈揺れとなったため、規則を定めた。図 3 に規則を示す。

上記の説明を行った後に、コーパスに利用するレビューとは別に、次節で述べる各製品ジャンルから 40 文ずつ、計 200 文のレビューを Amazon から収集し、コーパス作成に参加する注釈者に注釈付けの練習を行ってもらった。また練習の後、注釈間違いや注釈揺れについては指摘をし、注釈者間で話し合いを行い、意思の統一を試みた。

- ・タグ内文字列の末尾が句読点、感嘆詞、疑問符、助詞だった場合、それはタグの中を含めない
- ・「～ならば」のような条件部分はタグに含めない
- ・細かな説明を修飾する文節はタグの範囲に含めない。
- ・副詞的表現が直近にない場合はタグの範囲に含めない
- ・サ変動詞はタグの範囲に含める。
- ・value-attribute の順番で出現し、一語になっている部分はまとめて value とする。
- ・attribute と value となるべく細かく分解する。

図 2. 追加した規則

- ・「たぶん～」等、モダリティの呼応要素はタグの範囲に含めない。
- ・「～よ。」「～ね。」等、終助詞等はタグの範囲に含めない（伝達態度のモダリティ）
- ・「～でしょう」等、推定推量を表すモダリティ部分はタグの範囲に含めない（真偽判断のモダリティ）
- ・疑問部分には注釈付けを行わない。（表現タイプのモダリティ）

図 3. モダリティ部分に関する規則

3. 4 評判情報コーパスの作成

前述の規則を用いて、評判情報コーパスの作成を行った。

注釈付けに用いた文書は、Amazon から収集したレビュー 1 万文である。Amazon の製品分類を基にレビューを、電化製品、ホーム・キッチン、映像・音楽、ソフトウェア、ホビー・おもちゃの 5 ジャンルに分類し、それぞれから 2000 文ずつを収集した。

さらに、1 万文のうちの 1000 文は事前に第一著者が注釈付けを行い、事例用コーパスとして用いた。注釈者は当研究室の学生 10 名である。

注釈付けは 9000 文を 200 文程度に分割し、作業時間に余裕がある人に順次配布していく形で行った。作業量は注釈者により 2000 文から 200 文まで様々である。

最終結果として得られた、1 万文のコーパスに注釈付けされた各タグの数を表 1 に示す。なお、表 1 のタグ数は注釈付け作業後に、後で述べる入力ミス等の修正作業を行った後の値である。

表 1. 注釈付けされたタグ数

	タグ数	item	att	value	eval
電化製品	総数	1374	1077	1746	513
	異なり数	871	598	1407	331
ホーム・キッチン	総数	1077	875	1838	363
	異なり数	648	476	1523	242
映像・音楽	総数	1187	614	1026	367
	異なり数	694	378	871	253
ソフトウェア	総数	1243	821	1585	302
	異なり数	710	560	1316	209
ホビー・おもちゃ	総数	1247	785	1687	436
	異なり数	797	453	1441	285
全データ	総数	6128	4172	7882	1981
	異なり数	3522	2069	5821	971

3. 5 注釈付けの一致率

今回作成したコーパスの注釈付け精度を確認するために、注釈付けの一致度を調べた。

注釈付け作業終了後に、コーパス作成を行った注釈者の中から 4 名に、コーパス中の同じ 185 文に対して注釈付けを行ってもらった。この 185 文は 4 名の注釈者がはじめて見るレビューであり、内訳は、電化製品：37 文、ホーム・キッチン：35 文、映像・音楽：39 文、ソフトウェア：35 文、ホビー・おもちゃ：39 文である。185 文中でタグが付与された箇所 525 カ所について各注釈者間の一致率を κ 値で求めた。結果を表 2 に示す。

表 2. 注釈付けの一致率

注釈者の組	一致率 (κ 値)
注釈者 1・2	0.48
注釈者 1・3	0.54
注釈者 1・4	0.53
注釈者 2・3	0.41
注釈者 2・4	0.51
注釈者 3・4	0.48

κ 値は主観が入る判定が偶然に抛らず一致する割合であり、0.4~0.6 の間ならば中程度、0.8 を超えるとほぼ完璧な一致と考えられる。本実験では、全注釈者の組で中程度の一致といえる結果が得られた。

3. 6 注釈揺れの修正

10名の注釈者によるコーパス作成を行った後、注釈者の入力ミス、同一文字列に対する注釈揺れ等の明らかに分かる誤りに関しては修正作業を行った。

最初に注釈者の入力ミス、入力忘れが判明した部分を修正した。入力ミスが確認されたタグは全部で705個、コーパス作成作業で注釈付けされた全タグのうちの4.42%であった。入力ミスの内容としては、「タグに付与する属性情報の入力忘れ」「入力する属性項目が入れ替わってしまっている」「注釈付けツール使用時にマウスの操作を誤ったために注釈範囲が誤り謝っている」等が主なものであった。

以上の入力ミスは注釈付けツールのインターフェースに起因すると思われるものもあり、ユーザの使い勝手という面から注釈付けツールの改良を検討する必要がある。

次に、同一文字列の同一範囲に対し、異なるタグが注釈付けされている部分について、修正を行った。表3に注釈揺れの種類と数を示す。

表3. 同一文字列の同一範囲に対する注釈揺れ

注釈揺れの種類	修正前の個数	修正後の個数
item・att	76	11
item・val	16	15
item・eval	1	1
att・val	29	7
att・eval	6	0
val・eval	79	0
計	207	34

各注釈揺れを見直し、事前説明時の説明に沿わない部分、明らかに一部の注釈者のみが異なるタグを付与している部分については修正をし、一方のタグに揃えた。

同一文字列に対する異なるタグの中で、一方のタグに寄せることができなかった部分について述べる。itemに関連する部分では、製品の特徴が名称として用いられている場合が12個で最も多かった。例えば、「ベスト」は評価を表す表現と考えられるが、「ベストアルバム」という商品を呼称する際にも「ベスト」という表層が用いられる。同様の例として「コンパクト、ブラック」等が挙げられる。

また、視点となるitemの外部にあるitemはattributeとして扱われるために修正不可能なものが6個あった。同様の理由でitemとvalueが同じ表層になっているものが4個あった。例を図4に示す。

・ item-attribute の揺れ	
この PC	はメモリ の容量は 512MB です。
このソフト	はメモリ が 256MB 必要。
・ item-value の揺れ	
紙	の色が黄色いよ。
台座部分	の素材は紙だろう。

図4. 視点となるitemに起因する揺れの例文

attributeとvalueについて修正の後に残ったのは、サ変動詞の省略が行われているために表層が同じになっているもの、名詞と活用形が同じになってしまったもの、「～

的」という表現の部分である。他に、同音異義語、言葉の意味自体がよく分からないために修正すべきではないとした部分がある。

最後に、さらに追加で行った修正について述べる。今回のコーパス作成では、「itemタグやattributeタグのみが現れた場合には注釈付けを行わなくてもよい」と注釈者に説明したが、後述する抽出実験の予備実験を行ったところ、itemの抽出精度が著しく低い結果となった。これはitemタグが付与されている表層文字列が場所により注釈付けされていたり、いなかったりと学習がうまく進まなかったためだと考えられたため、itemに関しては全文を第一著者が見直し、追加で注釈付けを行った。修正前と修正後のitemタグの数と抽出精度を表4に示す。表4に示す抽出精度は、ジャンルごとにitemの自動抽出を行った際のF値の平均である。

表4. itemタグの修正前後の比較

	itemタグの数	抽出精度
修正前	2152	20.69
修正後	6128	44.88

4. 評判情報の抽出

今回作成したコーパスの利用として、評判情報の各構成要素について抽出実験を行った。抽出には固有表現抽出に用いられる手法を利用した。

抽出器にはCRF++[8]を利用し、作成したコーパスを製品のジャンルごとに分け、5分割交差検定を行った。なお、事前に半角英数字は全角英数字に直してある。また、形態素解析にはChasenを利用した。

素性展開については文字単位としたbaselineとして使用した素性は「表層文字、品詞、形態素原形、文節内素性、主辞素性」である。文節内素性とは、文節内の最も先頭に近い固有名詞の品詞細分類を、固有名詞がなければ文節の先頭の単語を素性として用いるものである。主辞素性とは連続する名詞では最後の名詞、連続していない名詞ではその名詞を素性として用いるものである。これらの各素性は固有表現抽出において効果が確認された素性である[3]。加えて、以下の①～③を素性とし抽出実験を行った。

①：角川類語新辞典の概念分類番号

②：活用形の原形

③：Jumanにより付与された単語の代表表記

①は表層表現が異なる意味的に近い語に対する素性、②は語尾の異なる同じ形態素に対する素性、③は表記の多様性に対する素性である。

抽出精度の結果を表5に示す。結果を見ると、itemの抽出における素性①③とevaluationの抽出における素性①②③についてはわずかながらに素性の有利不利の傾向を見て取ることができるが、全般的に大きく有効な素性を発見することはできなかった。一方タグ別に見ると、表層の種類が多様であるvalueの抽出精度が他の要素に比べて低くなっている。また、ジャンル別に見ると、映像・音楽のレビューに対する抽出精度が他ジャンルに比べて低くなっている。これは、製品ジャンルの性質上、製品の様態を表す表現が長く、遠回しになるためである。

今後は、今回の実験で傾向が見られたitemとevaluationをattributeとvalueの抽出に用いて段階的な抽出をおこなうなどの検討が必要であろう。

表 5. 各製品における評判情報の構成要素抽出精度 (F 値)

	item							
	baseline	base+①	base+②	base+③	base+①②	base+①③	base+②③	base+①②③
電化製品	0.520	0.513	0.522	0.520	0.519	0.525	0.519	0.519
ホーム・キッチン	0.513	0.514	0.510	0.517	0.510	0.514	0.521	0.522
映像・音楽	0.372	0.372	0.398	0.378	0.371	0.373	0.377	0.380
ソフトウェア	0.461	0.450	0.461	0.458	0.460	0.461	0.454	0.457
ホビー・おもちゃ	0.385	0.395	0.389	0.396	0.402	0.410	0.395	0.409
	attribute							
	baseline	base+①	base+②	base+③	base+①②	base+①③	base+②③	base+①②③
電化製品	0.512	0.511	0.514	0.509	0.503	0.510	0.505	0.499
ホーム・キッチン	0.487	0.494	0.485	0.483	0.492	0.488	0.479	0.494
映像・音楽	0.353	0.358	0.363	0.356	0.360	0.358	0.363	0.350
ソフトウェア	0.392	0.386	0.387	0.390	0.385	0.385	0.387	0.385
ホビー・おもちゃ	0.512	0.527	0.519	0.513	0.529	0.521	0.516	0.517
	value							
	baseline	base+①	base+②	base+③	base+①②	base+①③	base+②③	base+①②③
電化製品	0.380	0.378	0.370	0.378	0.375	0.377	0.373	0.376
ホーム・キッチン	0.403	0.416	0.417	0.409	0.410	0.409	0.420	0.413
映像・音楽	0.318	0.332	0.322	0.315	0.338	0.332	0.322	0.336
ソフトウェア	0.369	0.347	0.360	0.352	0.348	0.357	0.366	0.358
ホビー・おもちゃ	0.343	0.359	0.356	0.348	0.366	0.352	0.359	0.358
	evaluation							
	baseline	base+①	base+②	base+③	base+①②	base+①③	base+②③	base+①②③
電化製品	0.481	0.493	0.495	0.497	0.513	0.514	0.499	0.518
ホーム・キッチン	0.532	0.536	0.539	0.544	0.550	0.531	0.535	0.533
映像・音楽	0.505	0.522	0.501	0.515	0.522	0.523	0.507	0.535
ソフトウェア	0.492	0.524	0.494	0.519	0.529	0.536	0.511	0.540
ホビー・おもちゃ	0.519	0.533	0.531	0.522	0.522	0.529	0.517	0.505

5. 終わりに

本稿では、複数注釈者の人手による評判情報コーパスの作成と、評判情報の抽出実験を行った

1 万文のレビューに対して注釈付けを行い、20163 個のタグを注釈付けした。注釈者間の一致率についても κ 値で 0.4~0.5 を保つことができた。また、コーパス作成後の修正作業で明らかになった点からコーパス拡張の際に有用な知見が得られた。最後に、作成した評判情報コーパスを利用した抽出実験を行い、F 値で 0.32~0.55 の精度であった。

今後の課題について述べる。コーパス作成という点から見ると、注釈者への事前説明の違いが一致率に及ぼす影響を調査する必要がある。また、コーパスの規模が十分かどうかを調査する必要があると考える。

抽出という点では、ジャンルを横断した抽出実験を行うと同時に、学習に用いる素性の検討、他の学習手法の検討・比較が挙げられる。

参考文献

- [1] Peter D. 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)
 [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 回 年次大会発表論文集, pp.67-70 (2003)
 [3] 中野桂吾, 平井有三: 日本語固有表現抽出における文節情報の

- 利用, 情報処理学会論文誌 Vol.45, No3, pp.934-941 (2004)
 [4] 洪陽灼, 白井清昭: 対話行為タグ付きコーパスの作成支援, 言語処理学会第 11 回 年次大会発表論文集, pp.815-818 (2005)
 [5] 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, 情報処理学会研究報告 NL165-4, pp.21-28 (2005)
 [6] 小林のぞみ, 乾健太郎, 松本裕治, 立石健二, 福島俊一: 意見抽出のための評価表現の収集, 自然言語処理, Vol.12, No.2, pp.203-222, (2005)
 [7] 宮崎林太郎, 前田直人, 森辰則: 評判情報注釈付けタグセットの提案, 言語処理学会第 12 回 年次大会発表論文集, pp.240-231 (2006)
 [8] 鹿島久嗣, 坪井祐太, 工藤拓: 言語処理における識別モデルの発展-HMM から CRF まで-, 言語処理学会第 12 回 年次大会チュートリアル資料 (2006)
 [9] 高橋哲朗, 乾健太郎: アノテーションツール“Tagrin”の紹介, 言語処理学会第 12 回 年次大会発表論文集, pp.228-231 (2006)
 [10] 宮崎林太郎, 前田直人, 森辰則: 人手による評判情報注釈付けにおける揺れの分析と注釈付け支援ツール, 情報処理学会研究報告 NL176-21, pp.143-150 (2006)
 [11] 徳久雅人, 松浦大樹, 村上仁一, 池原悟: マンガにおける表情に着目した情緒注釈付き対話コーパスの作成, 電子情報通信学会 2008 年総合大会講演論文集, pp.”S-44”-“S-45” (2006)
 [12] 鍛冶伸裕, 喜連川優: 自動構築した評価文コーパスからの評価表現の獲得, 電子情報通信学会第 18 回 データ工学ワークショップ (2007)
 [13] 益岡隆志: モダリティの文法, くろしお出版 (1991)