

Web 文書集合からの意見情報抽出と着眼点に基づく要約生成

立石健二 福島俊一[†]

小林のぞみ 高橋哲朗 藤田篤 乾健太郎 松本裕治[‡]

あらまし

本稿では、Web 文書から意見を抽出し、それらをレーダーチャートの形式で要約/視覚化する意見抽出分類システムを提案する。Web の意見は、商品購入の際の情報収集、市場調査等のマーケティング、企業のリスク管理等、さまざまな目的での利用が考えられる。Web の意見の収集/分析に関する研究には 2 つの課題がある、対象とする Web 文書から意見に該当する箇所を抽出すること、抽出した意見を要約/視覚化することである。本システムは、この 2 つの課題を 3 つ組 {対象物, 属性, 評価} のモデルと情報抽出の手法を用いて解決する。本システムを車に関するレビューサイトの 100 記事を対象として評価したところ抽出精度が適合率 82%, 再現率 52%であり、システムが出力したレーダーチャートと人手で作成したレーダーチャートが類似することを確認した。

Web Opinion Extraction and Summarization Based on Viewpoints of products

Kenji Tateishi, Toshikazu Fukushima[†]

Nozomi Kobayashi, Tetsuro Takahashi, Atsushi Fujita, Kentaro Inui, Yuji Matsumoto[‡]

Abstract

This paper proposes an opinion extraction and classification system, which extracts people's opinions from Web documents and summarize/visualizes them in the form of "radar charts". People's opinions on the Internet are available for many purposes such as surveys before purchasing products, market research and risk management for enterprises. There are two issues on this area. One is to locate opinion sentences from Web documents, and the other is to summarize/visualize the extracted opinions. The proposed system solves them by employing an opinion model {object name, attribute expression, evaluative expression} and information extraction techniques. The experimental result conducted with 100 articles on the car domain showed that the system performed 82% on precision and 52% on recall, and that both radar charts created by the system and by the hand are similar to each other.

1. はじめに

近年、インターネットの爆発的な広がりの中で、Web ページでは個人が自由に情報を公開できるようになり、掲示板や Web ログのようなコミュニティではさまざまな人の中での情報の交換が容易になった。このようにインターネットは、誰もが情報を発信できる場でありそこにはさまざまな人の多様な意見が存在すると考えられる。もし、これらの意見を効率的かつ安価に収集・分析できる方

法があれば、商品購入の際の情報収集、市場調査等のマーケティング、企業のリスク管理等、幅広い利用が考えられる[14]。

このような期待のもとに筆者らはこれまで、掲示板等のWebサイトから意見を収集・分析する研究を進めてきた[1, 13, 14]。また、我々の研究と並行して、意見に着目した研究が行われている[4, 5, 6, 7, 8, 9]。最近では、WebFountain[11, 12]のように数千台のPCを用いて収集した数十億のWebログや掲示板等から製品の評判をトラッキングするツールを開発するような、大規模な人的/資源的リソースを導入したプロジェクトが開始されている。

意見を収集・分析する研究には 2 つの課題がある。1 つ目は、対象となる Web 文書から意見が記述されている箇所を特定することである。掲示板

[†] NEC インターネットシステム研究所
Internet Systems Research Laboratories, NEC Corp.

[‡] 奈良先端科学技術大学院大学
Nara Institute of Science and Technology

等の Web 文書は、アンケート等と異なりその記述形式に自由度が高いため、そのままテキストマイニングツール等で分析することは難しい。例えば、「始めまして。昨日購入してまだいろいろ試しているところですが、ちょっと使った感じではキーボードがすごく良いですね。」では、記述者は実質的には「キーボードがすごく良い」ということを言っているに過ぎない。

2つ目の課題は、収集した意見を要約/視覚化することである。収集した意見をそのまま羅列しても、そこから意見の全体像を把握するのは難しい。また、意見の要約/視覚化は、競合商品間で意見を比較する場合や、分析の最初の起点を得るために重要である。商品のための意見分析では、注目する商品と比較対象の同一な点と異なる点を明らかにした上で、異なる点を詳細に分析していくという手順が有効だからである。

本稿では、上記の2つの課題を解決する意見抽出分類システムを提案する。本システムの動作イメージを図1に示す。本システムは、Web 文書から3節で定義する{対象物, 属性表現, 評価表現}の3つ組を抽出し、それらを集計してレーダーチャートを作成する。意見は、4節で示すように情報抽出の手法に沿って辞書と抽出ルールを用いて抽出する。抽出した意見は、属性表現, 評価表現を手がかりとして着眼点, 評価値(肯定/否定)の軸へ分類し、レーダーチャートを作成する。5節の分析例から、比較分析が容易になること、分析の起点にできることを示す。また6節の評価実験では、車に関するレビューサイトの100記事を対象として評価したところ、本システムによる抽出精度が適合率82%、再現率52%であり、システムが出力したレーダーチャートと人手で作成したレーダーチャートが比較的一致することを確認する。

2. 機能要件

最初に、開発する意見抽出分類システムに必要な機能を説明する。意見の分類軸として、着眼点(評価の観点、と評価値(肯定/否定)の2つを考える。この場合、本システムを実現するためには、下記の3つの機能が要件となる。

- (1) Web 文書から意見に該当する箇所を抽出する機能
- (2) 抽出した意見を着眼点の軸で分類する機能
- (3) 抽出した意見を評価値(肯定/否定)の軸へ分

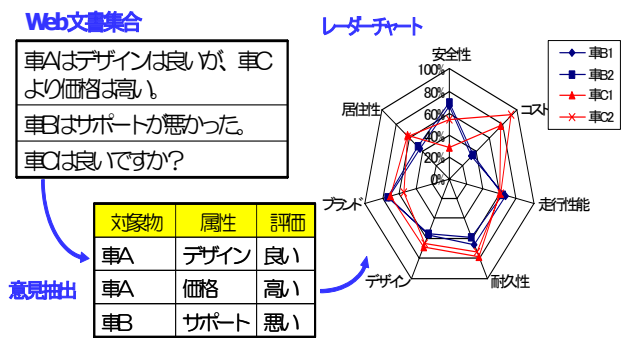


図1 システムの動作イメージ

類する機能

これらの機能が実現できれば、図1のようなレーダーチャートの作成が可能である。図1の各軸は「安全性」「コスト」「走行性能」等の意見の着眼点に対応する。軸の値(評価スコア)はその着眼点に属する意見の内の肯定意見の割合である。

3. アプローチ

3. 1 意見のモデル

2節の要件を満たす意見抽出分類システムを開発するために、情報抽出の技術を用いる。まず意見を下記の2つの条件を満たす情報と定義する。下記の3つ組の意見のモデルは[1][2]において提案したものである。

- (i) 表1の{対象物, 属性表現, 評価表現}の3つ組で構成される情報である。
- (ii) (i)の3つ組には記述者の判断として3つの構成要素間の関係が存在する。

ここで、それぞれの条件について詳しく説明する。まず、(i)の条件について、表1にあるように対象物とは、例えば、商品名、企業名、人名が該当する。次に、属性表現とは、**対象物の特徴や性質を示す表現**であり、例えば、コンピュータの分野では、デザイン、性能、価格、サポートが該当する。評価表現とは、**肯定又は否定の評価を示す表現**であり、例えば、良い、悪い、使いやすい該当する。

次に、(ii)の条件について、表2の(a)-(e)の文は(i)の条件を満たす文、すなわち対象物, 属性表現, 評価表現を含む文である。このうち、(a)は(ii)の条件も満たすため、そこから{Mobile777, デザイン, 良い}という3つ組を意見として抽出する。一方、(b)-(e)の文は(2)の条件

表1 意見のモデル

種類	説明	例
対象物	商品, 企業名, 人名等	LaVie, NEC, アインシュタイン
属性表現	対象物の特徴・性質、対象物の部分	性能, 価格, デザイン, サポート
評価表現	肯定又は否定の評価	良い, 好き, 速い, 使いやすい

表2 意見の定義を満足するケースと満足しないケース

例文	意見
(a) <u>Mobile777</u> の <u>デザイン</u> が <u>良い</u> 。	○ {Mobile777, デザイン, 良い}
(b) ... <u>デザイン</u> が <u>良くて</u> , <u>Mobile777</u> の <u>性能</u> が.....	× (3つの関係がない)
(c) <u>Mobile777</u> の <u>デザイン</u> が <u>良いの?</u>	× (疑問)
(d) <u>Mobile777</u> の <u>デザイン</u> が <u>良いと聞いた</u> 。	× (伝聞)
(e) <u>Mobile777</u> の <u>デザイン</u> が <u>良ければ買う</u> 。	× (条件)

を満たさないため、3 つ組を抽出しない。(b)は3つの表現間の関係がないからである。(c)-(e)は3つの表現間の関係はあるものの、(c)は疑問、(d)は伝聞、(e)は条件を示し、記述者の判断を示さないためである

3. 2 2節の機能の実現

上記の意見の定義を導入することにより、2節の機能は、次のような方式により実現できる。まず、要件(1)は3.1節の定義を満たす3つ組を意見として抽出することにより実現する。要件(2)の着眼点への分類は、抽出した3つ組の意見の属性表現を手がかりとして分類する。要件(3)の評価値への分類は、抽出した3つ組の意見の評価表現を手がかりとして分類する。

ただし、要件(3)の評価値への分類は、評価表現単独では決定できない評価表現がある。例えば、「良い」「好き」については、これら単独の評価表現から評価値を求めるが、「高い」の場合「価格が高い」と「性能が高い」のように「高い」だけでは評価値を求めることが出来ない。このような一部の評価表現については、属性表現と評価表現のペアから評価値を求める。

4. システム構成

本システムは、情報抽出の手法に沿って、辞書と抽出ルールを用いて意見を抽出・分類する。{対象物・属性表現・評価表現}辞書は、あらかじめ作成しておく。特に属性表現・評価表現辞書は、[2]で提案した方式を利用して対象物の分野毎に半自動的に作成する(4.1節参照)。システムは、まずこれらの3種類の辞書と抽出ルールを用いて意見を抽出する(4.2節参照)。次に、属性・評価表現辞書に付与した分類ラベルを参照して意見を着眼点と評価値の軸で分類し(4.3節参照)、レーダーチャートを作成する。

本システムによる意見の抽出分類の手順を図2の例を用いて説明する。図2では、入力として2つの文が与えられている。まず、システムこれらの文から{車A, 燃費, 良い} {車B, 加速, 鈍い}の3つ組を抽出する。対象物・属性・評価表現辞書に登録されているそれぞれの辞書表現と、3表現がまとまりでありかつ記述者本人の判断であることを示す抽出ルールが存在するため抽出できる。次に、属性表現辞書に登録した「燃費」「加速」に対するラベルを用いて抽出した意見を「コスト」「走行性能」の着眼点に分類する。同様に評価表現辞書に登録した表現のラベルを用いて抽出意見を「肯定」「否定」の評価値に分類する。

4. 1 属性・評価表現辞書作成

辞書作成支援ツールは、属性・評価表現辞書として小規模の初期辞書を用意しておけば、共起パターンを用いて大量文書集合からブートストラッピング的に双方の辞書を交互に増やすことができるツールである。ユーザは、ツールが提示した表現候補の採否を入力するだけで半自動的に大規模な辞書作成を効率的に実施できる[2]。

ここで、このブートストラッピング方式について図3の例を用いて簡単に説明する。まず、ユーザは初期の対象物と評価表現の辞書を用意する。この2つの辞書と共起パターンを用いて、ツールは「対象の属性はX」のXに当てはまる様々な表現を特定の順番で評価表現候補として出力する。ユーザはそれらの評価表現候補の採否を決定する。ツールはユーザが採用した表現を評価表現辞書に登録する。次に、この評価表現辞書と最初の対象物の辞書を用いて、ツールは「対象のYは評価」

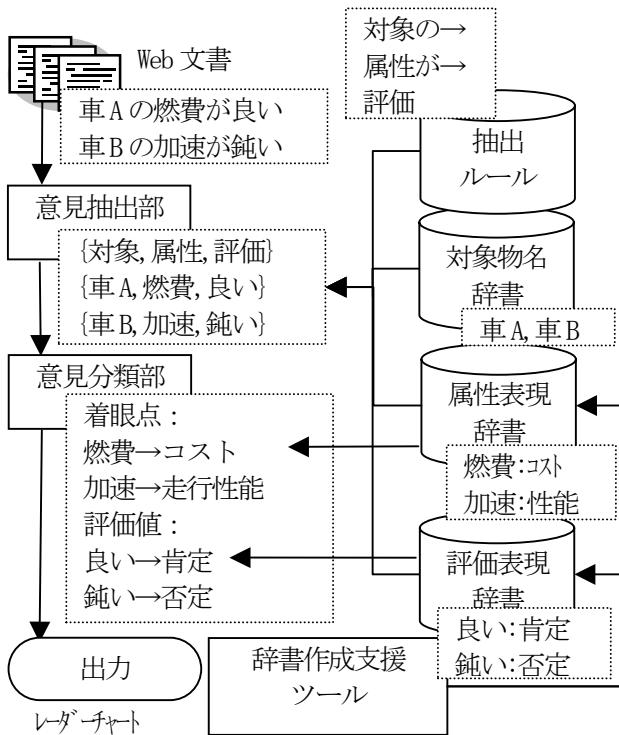


図2 システムの構成

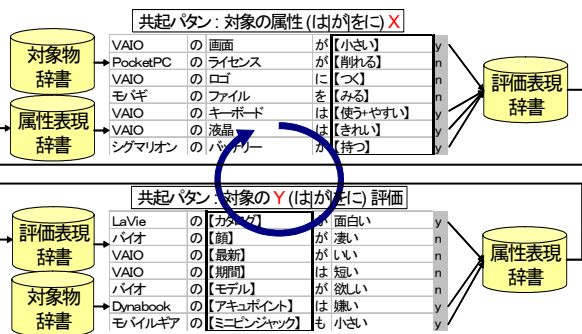


図3 辞書作成支援ツールの動作イメージ

の Y に当てはまる様々な表現を特定の順番で属性表現候補として出力する。ユーザはそれらの属性表現候補の採否を決定する。ツールはユーザが採用した表現を属性表現辞書に登録する。このサイクルを繰り返すことにより、2つの辞書を少しずつ増やしていくことが可能である。

このような方式で収集する属性・評価表現に対して、着眼点・評価値の2種類のラベルを手で付与する。着眼点の種類はユーザがその用途に合

図4に辞書作成支援ツールの画面例を示す。3番目の列に属性表現の候補を表示しており、利用者は一番左側の列に表現の採否を入力する。表現を採用する場合は着眼点のラベルを、不採用の場合は“n”を入力する。表現の収集と同時にラベルを付与する。それぞれの属性表現の上をクリック

表3 作成した辞書の例

着眼点	属性表現の例
安全性(47)	ABS, エアバッグ, セキュリティ
コスト(52)	価格, 維持費, 燃費, 修理代
サービス(35)	アフターサービス, 営業, 対応
走行性能(411)	エンジン, ギア, 立ち上がり
耐久性(83)	ガタツキ, キズ, ボディ剛性
デザイン(155)	スタイル, 外観, 格好, 形状
ブランド(147)	販売戦略, CM, ネームバリュー
メンテ(57)	アフターパーツ, オイル, 交換
居住性(381)	エアコン, オデレ, 音, 荷物
評価値	評価表現の例
肯定(728)	いい, エレガント, おしゃれ
中立(16)	普通, まあまあ, 一長一短, 並
否定(619)	イマイチ, 嫌い, うるさい

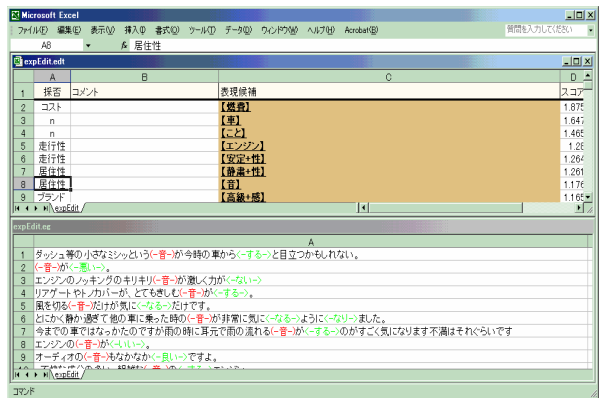


図4 辞書作成支援ツールの画面例

すると例文が表示するため、実例を見ながら採否の判断が可能である。

4.2 意見抽出方式

意見抽出は3つの手順で進める。まず対象物・属性・評価表現の文書内の位置をそれぞれの辞書を参照して検出する。

次に、評価表現を中心として関係する属性表現・対象物を検出し、対象物・属性・評価の3組を抽出する。この処理は、属性→評価間の関係と対象物→評価間の関係を検出する処理に分けて考える。どちらも、主語→述語の関係であり、同一の枠組みが適用可能だからである。また、表4の係り受け関係を利用した抽出ルールを用いる。抽出ルールは、辞書作成支援ツールで用いる共起パターンに類似するが、再現率を高めるため、より広い範囲の複数の文節に跨る関係も記述している。これらの抽出ルールのいずれかに適合する場合は

表4 抽出ルール(属性:属性表現, 評価:評価表現, 属性を対象物に置き換えたルールも用意)

抽出ルール	例
(属性:が は も の に を で φ)→(評価)	<u>デザインが</u> 良い, <u>外観</u> 良い
(評価:<連体修飾>)→(属性)	良い <u>デザイン</u>
(属性)=(評価)	<u>デザイン</u> グッド
(属性:の)→(*:が は も)→(評価)	<u>デザイン</u> の質が 良い
(属性:も や と 、)→ ([*]: も は が で)→(評価)	<u>デザイン</u> も広告 も良い

表5 意見性判定ルールの例(評価:評価表現)

意見性判定ルール	意味
(1) 評価.*か(? ?) 評価.{0,8}(ろうか のか るか たか)	疑問文 (～良いでしょうか?)
(2) 評価.{0,12}(人は 人も 人には ...)	条件1 (～が良い人には)
(3) 評価.{0,16}(そうだ 言われ 聞いた いわれ)	伝聞 (～良いらしい)

両者に関係があるとする。

最後に、抽出した3つ組の意見性を判定する。抽出した3つ組の中には表2の(c)-(e)のように、記述者の判断を示さないものも存在する。そこで、上記の抽出ルールとは別に評価表現の近傍の条件を示す接続詞、及び質問・伝聞等を示す文末表現等を考慮した意見性判定ルールを用意し、それらに該当する場合は意見でないとして除外する。使用した意見性判定ルールの一部を表5に示す。

4. 3 意見分類方式

原則として抽出した意見に付与した着眼点・評価値のラベルに従う。評価値の分類に関して、評価表現の近傍に奇数回の否定表現(例. ない)が存在する場合には評価値を反転する。

例) Mobile777は良くない。→否定(奇数回)

例) Mobile777は良くなくない。→肯定(偶数回)

5. 分析例

4節の方式を用いて作成したレーダーチャートの例を図5にしめす。このチャートは、車に関するレビューサイトの630記事から意見を抽出したものである。(a)のチャートは同車種間の比較であ

り、(b)のチャートは異車種間の比較である。(a)では、同車種であるためチャートの形状は類似していることがわかる。多くの着眼点の軸でA2よりA1の方が利用者の評価が高い。一方、(b)では4種類の車を比較しているが、B1とB2、C1とC2は同系列の車であるためやはり類似したチャートの形状を持つ。B系列の車は安全性と走行性能、ブランドイメージにおいてC系列の車よりも評価が高い。C系列は、コストが特に良くC2においてはほぼ満点である。このように、意見をレーダーチャートという形式で要約することで商品の比較が容易になる。

次に、(a)のチャートのA1とA2の意見の相違点を詳しく分析する。表6は、走行性能に関する意見を出現頻度順に並べたものである。A2は、「ナビ」「シフトショック」「装備」の否定意見が上位にあるため、ここが評価の差となったのではないかと読み取れる。このように、レーダーチャートによる要約を起点として、ポイントを絞った詳細な分析が可能となる。

チャートのために使用した辞書は、同サイトの1000記事から約5時間で作成した。1445表現の属性表現辞書と1363表現の評価表現辞書である。評価表現辞書のうち53表現は属性表現と評価表現の組に対して評価値を付与した。作成した辞書の例は、表3に示した通りである。属性表現に対しては9つの着眼点のラベルを用意したが、チャート作成に用いたのは意見の出現頻度が10未満の2つを除いた図5の7つの着眼点である。なお、本システムの開発において、構文解析ツールにはCaboCha[3]を使用した。

6. 評価

6. 1 抽出精度評価

まず、人手で付与した三つ組のタグを正解事例として、4節で説明した意見の抽出精度の実験を行う。

実験対象のデータは、車に関するレビュー記事を100記事選択した。これは、5節の分析例で用いたのと同サイトの記事である。次に、それぞれの記事に対して3つ組のタグを人手で付与した。実際にタグを付与してみると、対象物名と属性表現が省略される場合があり、表7にあるようにタグには4種類がある。本実験では(1)に加え(3)も

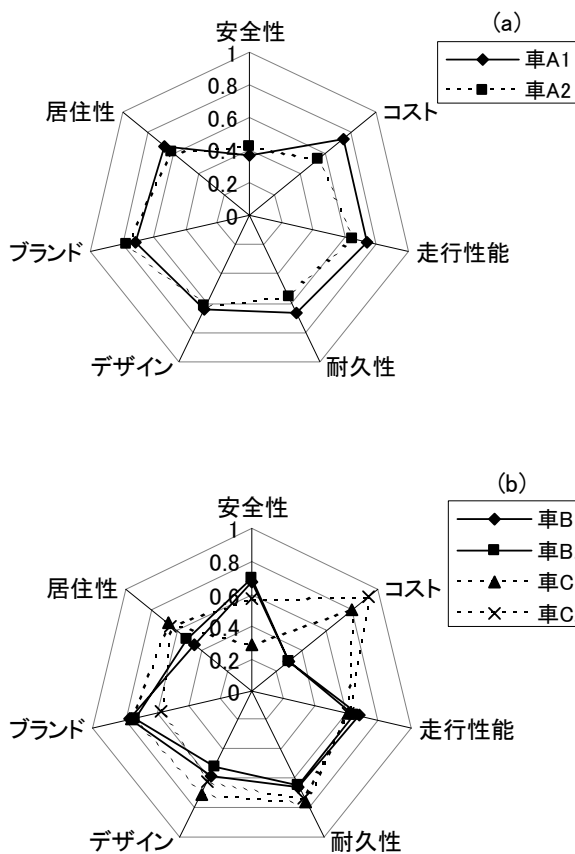


図5 レーダーチャート作成の例

抽出対象とし、母集団は、(1)+(3)の 36+1007=1043 件とした。その理由は、掲示板の場合は本文中に対象物名が記述されていなくとも記事タイトルに対象物名が記載されており、省略された対象物名を補充することができるからである。

実験には 2 種類の辞書を用いた。1 つ目は、同一サイトの別の 1000 記事から作成した対象、属性、評価表現辞書である（別記事の辞書）。これは 5 節の分析例で用いたものと同一である。2 つ目は、実験対象と同じ 100 記事から作成した（実験記事の辞書）。つまり、100 記事に付与された対象、属性、評価タグをそのまま対象、属性、評価表現辞書として登録した。これによって、辞書の大きさの影響を含めない純粋な抽出精度も評価する。

実験結果を表 8 に示す。別記事の辞書の場合には、辞書の不備による影響で、適合率/再現率共に実験記事の辞書よりも 10%程度低くなっている。不正解の原因を実験記事の辞書の場合で分析すると、まず、抽出誤りについては、732-677=55 件の誤りのうちの約 7 割が抽出した組が意見ではない

表6 着眼点「走行性能」に関する意見の例（左：車A1 右：車A2）

属性表現	評価	数	属性表現	評価	数
エンジン	肯定	25	エンジン	肯定	21
トルク	肯定	13	運転	肯定	12
走り	肯定	12	走り	肯定	7
パドルギ	肯定	8	加速	肯定	7
パワー	肯定	8	ナビ	否定	6
加速	肯定	7	足回り	肯定	5
運転	肯定	7	パワー	肯定	5
走行性能	肯定	5	シフトショック	否定	4
バランス	肯定	5	装備	肯定	4
装備	肯定	4	装備	否定	4
回転	肯定	4	ナビ	肯定	4

場合であった。4.2 節において意見性判定ルールを用いて意見性を判定しているものの、それには当てはまらないケースが出現したためである。次に、抽出漏れについては、1043-677=366 件のうち約 8 割が抽出ルールの不一致により抽出できない場合であった。4.2 節の表 4 の抽出ルールを用いているものの、それらに当てはまらない属性と評価、対象物と評価の関係が出現したためである。両者のケースとも、今後のルールを補充が課題となる。その他の原因としては、抽出誤りについては構文解析誤りによるもの、抽出漏れについては評価表現が複数文節に跨り本システムでは抽出できない場合が多かった。

6. 2 主観評価

次に、人手で作成した特定商品に関するレーダーチャートと、4 節の方式によって作成したレーダーチャートを比較する。実験には、5 節の分析例で用いた 4 商品(A1, A2, B2, C2)を用いた。それぞれの商品は 50-100 記事からなる。システム側のレーダーチャートは、5 節と同一のものである。

人手によるレーダーチャートは、5 人の評価者によって作成した。評価者は、各商品毎にシステム側が用いたのと同じの記事群を一度に読み、その後 7 つの着眼点の軸それぞれについて 11 段階(0-1.0, 0.1 刻み)の評価スコアを付与した。その着眼点の軸に関して肯定的であるほど高い評価スコアを、否定的であるほど低い評価スコアを与えた。この作業を 4 商品について実施した。最終的に 5 人の評価スコアの着眼点の軸毎の平均値を求

表7 実験記事に付与したタグの種類

タグの種類	頻度(計 1171)
(1) {対象, 属性, 評価}	36
(2) {対象, , 評価}	23
(3) { , 属性, 評価}	1007
(4) { , , 評価}	105

表8 意見抽出精度(実験記事:実験記事から辞書を作成, 別記事:別記事から辞書を作成)

辞書の種類	適合率	再現率
実験記事	92%(677/732)	65%(677/1043)
別記事	82%(547/670)	52%(547/1043)

め、レーダーチャートを作成した。平均値を求める際に評価者毎のスコアの正規化は行っていない。

なお、評価を客観的なものにするために、記事内の商品名およびメーカー名は伏せ字にした。また、評価者は全員 20-40 代の女性であり、自動車免許を保有しているが、車を取り扱うような職業を有する人物ではない。さらに、評価者には記事を読む前に、後で着眼点の軸について評価スコアをつけてもらうことをあらかじめ伝えた。

本システムの着眼点の軸の値と人手の評価スコアの値の着眼点の軸毎の差を求めた。7 着眼点軸 x 4 商品の計 28 の着眼点の軸のなかで 10 軸が 0.05 以内, 18 軸が 0.1 以内, 23 軸が 0.15 以内の差であった。例えば図 6 は、A1 と B2 に関する両者のチャートを比較したものであるが一部の着眼点の軸を除き両者の値は近いところにマッピングされている。そのため、本方式により作成したレーダーチャートは人が作成した場合に比較的近いものが得られるといえる。

7. 関連研究

本研究では、情報抽出のアプローチを用いて、{対象物, 属性, 評価}の 3 つ組の意見を抽出分類し、レーダーチャートという形式で意見を要約することを目的としている。

近年、意見の分類に関する研究が盛んである。この分野の研究は、(a)意見を肯定/否定に分類する研究[4, 5], (b)主観的な文と客観的な文に分類する研究[6, 7], (c)感情のカテゴリに分類する研究[8], (d)発話意図のカテゴリに分類する研究[9], がある。(a)では、レビューサイトの記事とそれらのレーティング情報を学習データとして、テキス

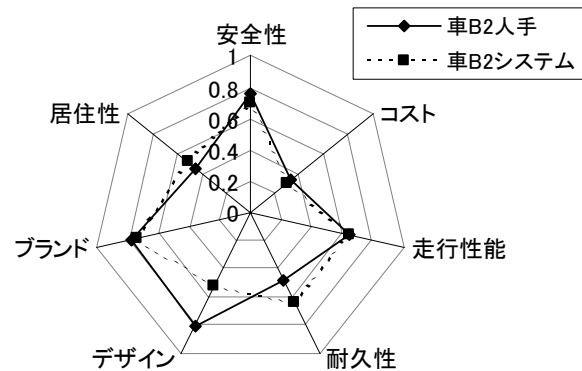
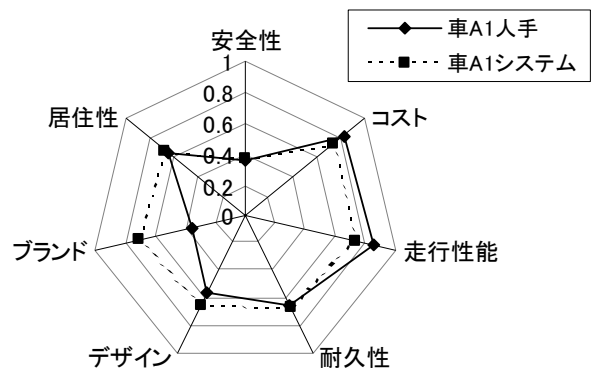


図6 レーダーチャートの比較

ト分類の手法により文を肯定・否定に分類している。(b)では、新聞記事を対象として、主観的な意見が記述された文書(例. 社説)と、客観的な事実が記述された文書(例. ニュース記事)をテキスト分類の手法により分類する。(c)では、電子メールの文を、知識ベースを用いて6種類の基本的な感情のカテゴリ(happy, sad, anger, fear, disgust, surprise)に分類する。(d)では、アンケートの自由回答文を対象として、文末表現を用いながら、文を質問、回答、依頼等の発話意図に分類する。

しかし、いずれの研究も、{対象, 属性, 評価}という意見の構成要素を取り扱うわけではないので、どの商品に対する意見であるか、どの着眼点に関する意見であるかまでを知ることはできない。そのため、レーダーチャートを作成するという研究目的達成には不十分である。

一方、情報抽出の研究領域では、意見の構成要素を抽出するという着眼点での研究はなかった。我々は今回、辞書と抽出ルールに基づく方式によって意見抽出を実現した。これは固有表現抽出の分野ではしばしば用いられる枠組みである。固有

表現の抽出ルールを自動作成する研究[10]、固有表現辞書をブートストラッピングによって自動作成する手法[15]を初めとして、今後方式を適用可能であると考ええる。

8. おわりに

本稿では、Web 文書から{対象物・属性・評価}の3つ組の意見を抽出・分類し、レーダーチャートを作成する意見抽出分類システムを提案した。本システムでは、情報抽出技術を用いて、辞書と抽出ルールを用いて意見を抽出した。実際の分析例から、比較分析が容易になること、分析の起点にできることがわかった。また評価実験では、本システムを車に関するレビューサイトの100記事を対象として評価したところ、本システムによる抽出精度が適合率82%、再現率52%であり、システムが出力したレーダーチャートと人手で作成したレーダーチャートが比較的一致することを確認した。

今後の予定として、意見抽出精度の向上がある。4.2節で用いた抽出ルールおよび意見性判定ルールを効率的に増やす方法を検討する。また、辞書作成の効率化として、ブートストラッピング的な手法で辞書を自動作成する方式[7]を応用して収集時間の短縮化を図る予定である。

参考文献

- [1] Tateishi, K., Ishiguro, Y., Fukushima, T., "A Reputation Search Engine that Collects People's Opinions by Information Extraction Technology", IPSJ Transactions on Databases, Vol.22, 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, NL154-012, pp. 77-84, 2003.
- [3] 工藤拓 松本裕治, "チャンキングの段階適用による日本語係り受け解析", 情報処理学会論文誌, Vol.43, No.6, pp.1834-1842, 2002.
- [4] Pang, B., Lee, L., and Vaithyanathan, S., "Thumbs up? Sentiment classification using machine learning techniques", Proceedings of the Conference on Empirical Methods in Natural Language processing (EMNLP2002), pp.77-86, 2002.
- [5] Kushal Dave, Steve Lawrence, and David M. Pennock, "Mining the peanut gallery: Opinion Extraction and Semantic Classification of Product Reviews", WWW2003, 2003.
- [6] Yu, H., Hatzivassiloglou, V., "Towards Answering Opinion Questions: Separating Facts from Opinions and Identifying the Polarity of Opinion Sentences", 2003 Conference on Empirical Methods in Natural Language Processing (EMNLP2003), 2003.
- [7] Riloff, E., and Wiebe, J., "Learning extraction patterns for subjective expressions", 2003 Conference on Empirical Methods in Natural Language Processing (EMNLP2003), 2003.
- [8] Liu, H., Lieberman, H., and Selker, T., "A Model of Textual Affect Sensing using Real-World Knowledge", Proceedings of the 2003 International Conference on Intelligent User Interfaces (IUI2003), 2003
- [9] 乾裕子, 内元清貴, 村田真樹, 井佐原均, "文末表現に着目した自由回答アンケートの分類", 情報処理学会自然言語処理研究会, 98-NL-128, p.181-188, 1998
- [10] Kiyoshi Sudo, Satoshi Sekine, and Ralph Grishman, "An Improved Extraction Pattern Representation Model for Automatic IE Pattern Acquisition", ACL2003, pp. 224-231, 2003.
- [11] <http://www.almaden.ibm.com/webfountain/>
- [12] Gruhl, D., Chavet, L., Gibson, D., Meyer, J., Pattanayak, P., Tomkins, A., Zien, J., "How to build a WebFountain: An architecture for very large-scale text analytics", IBM System Journal, Vol. 43, No. 1, 2004.
- [13] Morinaga, S., Yamanishi, K., Tateishi, K., Fukushima, T., "Mining product reputations on the Web", KDD2002, pp 341-349, 2002
- [14] 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, Vol. 19, No. 3, pp. 317-323, 2004.
- [15] Riloff, E., and Jones, R., "Learning Dictionaries for Information Extraction by Multi-level Bootstrapping", In Proceedings of the Sixteenth National Conference on Artificial Intelligence(AAAI-99), 1999.