

意見分析コーパスの現状と課題

関 洋平^{1,a)}

受付日 2013年3月20日, 採録日 2013年7月5日

概要: ネットメディアの発展とともに, ユーザ個人が発信する意見の影響は大きくなり, その利用方法は変遷し続けている. オンライン上の意見情報を分析することにより, 消費活動や行動判断に有用な情報へのアクセスを支援する研究として, 主に自然言語処理, 情報検索, データ工学などの分野で, 意見分析あるいは意見マイニングの研究が行われている. 意見を分析するための表層的な特徴は, 文書ジャンルあるいはメディアごとに異なることが知られており, ニュース, レビュー, ブログ, マイクロブログなど, さまざまな文書ジャンルごとにコーパスが重要な役割を果たす. 本稿では, 各文書ジャンルに代表的な意見分析コーパスの構築方針について説明し, 最新の意見分析研究における, コーパスの活用事例や, 従来の枠にとどまらない新たな取り組みについて, 文書ジャンルごとに整理する. また, コーパス構築の課題を解消するための取り組みとその適用の基準について議論する.

キーワード: 意見分析コーパス, 文書ジャンル, コーパスアノテーション

Current Status and Issues on Sentiment Analysis Corpora

YOHEI SEKI^{1,a)}

Received: March 20, 2013, Accepted: July 5, 2013

Abstract: With the rapid growing of internet media, consumer generated opinions are referred by many users and also utilized for variety of applications nowadays. To access of the valuable online sentiments and support the consuming or political decision, opinion mining and sentiment analysis researches are conducted in NLP, IR or DE fields. Surface sentiment expressions are described in a different style according to document genre or media, so developing sentiment corpus in each media, such as news, reviews, blogs, or microblogs is valuable and important. In this paper, we explain the annotation policy of representative sentiment corpora according to the document genre and summarize recent direction in researches using sentiment analysis corpora to clarify new annotation features and corpus usage methods. We also discuss criteria to solve annotation efforts in creating sentiment corpora.

Keywords: sentiment analysis corpora, document genre, corpus annotation

1. はじめに

ここ数年, Web などの大量の電子化テキストに現れる他者が発信した意見情報を抽出し, 集約や可視化を行うことで, 世論調査や評判分析といった応用を実現する研究が進んでいる [1], [2], [3], [4]. これらの研究を総称して, 意見分析 (*Sentiment Analysis*) あるいは意見マイニング (*Opinion Mining*) と呼ぶ [1]. 対象となる文書ジャンルは, 報道機

関が配信するニュース, Web 上のレビューサイト, 個人が自身の体験や意見を記述するブログやマイクロブログなどであり, 政策や選挙のための情報分析, 世論調査, 商品や映画やレストラン・ホテルなどのサービスの評判分析, トレンド分析, などについて実用化が進められている.

意見を分析するためには, 意見とは何かを定式化する必要がある. ただし, 一口に意見といってもその特徴はさまざまである. 意見の定義の範囲は広く, 主観性などの広い概念を対象とした場合, 評価, 感情, 意見, 態度, 推測などの何を対象とするかを決定することも重要である [5], [6]. 一方で, 文書ジャンルとは, 文書の書き手と読み手との間

¹ 筑波大学図書館情報メディア系
Faculty of Library, Information and Media Science, University of Tsukuba, Tsukuba, Ibaraki 305-8550, Japan

^{a)} yohei@slis.tsukuba.ac.jp

で、読む行為を通じたコミュニケーションの共通パターンを想定できる文書群を指す概念と位置づけることができる [7]。この考えに従うと、文書ジャンルが異なれば、文書中に現れる意見の伝達方法も異なるといえる。

次に、ニュース、レビュー、ブログやマイクロブログの文書ジャンルを利用した意見分析のタスクについて考える。

ニュースを対象とした意見分析では、さまざまな識者や組織により表明されている意見を話題別に集約するタスク [8] や、ニュースを発信するメディアの立場を把握するために、意見の差異を判別するタスクが考えられる。また、レビューなどと比べると、話題の範囲も広い。このため、ニュースを対象とした意見分析コーパスでは、肯定（ポジティブ）や否定（ネガティブ）といった観点（以降、**極性**と呼ぶ）に加えて、極性の度合い（強い肯定や弱い否定など、以降、**極性の強さ**と呼ぶ）、誰が意見を表明しているか（以降、**意見保有者**と呼ぶ）、何についての意見を表明しているか（以降、**意見対象**と呼ぶ）、どのフレーズが意見を表しているか（以降、**意見表現**と呼ぶ）といった情報の付与が必要となる。次の例文について、具体的に説明する。

- 「潘基文国連事務総長は、日本維新の会の橋下徹共同代表の発言について激しく批判した。」

極性とその強さは、強い否定（ネガティブ）、意見保有者は「潘基文国連事務総長」、意見対象は「橋下徹共同代表の発言」、意見表現は「激しく批判した」となる。ただし、意見保有者、意見対象、意見表現は、文字列の切り出しの範囲の定義に難しさがある。

一方で、極性の区別は、レビューなどの文書ジャンルと比べて、明確でなく、さまざまな課題が存在する。たとえば、以下の例文は、発言者の前向き姿勢を示す肯定的な意見か、現状の懸念という否定的な意見かは、一概に判断することは難しく、人によらず判定させるためには、何らかの判定基準が必要となる。

- 「東アジア諸国は、過去の不幸な苦難を乗り越える真剣な努力をしなければならぬ」と東アジア地域が不安定化することへの懸念を示した。

こうした問題に対応した意見分析コーパスを構築するために、複数のアノテータによる判定のゆれについて、個別の判定者のバイアスを修正し、安定的な判定基準を設定する取り組みが行われている [9]。

また、ニュース報道の特徴として、「わたし（一人称）の意見を表明していることを、他の文書ジャンルと比較して、明示的に記述しない」[10] 傾向があることが指摘されている。したがって、メディアの主張は、立場の近い識者の意見や、関連する事実を引用することにより行われることが多い。次の文について考える。

- 「ウミガメの保護観察など、復興予算の雇用対策の97%は被災者以外で使われている。」

これは関連する事実を通して、著者による復興予算の使

われ方についての否定的な意見を示唆しており、後の文では「大切な雇用でも復興予算のずさんな使われ方が続いている」と続く。こうした意見の分析には、談話全体の構造などを考慮する必要がある。

次に、レビューを対象とした意見分析では、商品やサービスの顧客を対象として、たくさんのレビューを読むことなく、購入の判断になる材料を提供するタスクや、業者を対象として、自社の商品についての評判や改善事項を確認するタスクが考えられる [11]。レビュー文書の特徴としては、特定の製品やサービスを対象としていることから、話題がある程度絞り込まれていることや、Web 特有のインフォーマルな表現の使用があげられる。また、タスクの応用目的から、単に商品の評判（肯定・否定）を判断するだけでは、顧客や自社のニーズに対して不十分であり、より細かな商品の特徴（以降、**アスペクト**と呼ぶ）を抽出し、アスペクトに対する評判を判別する研究がさかんに行われている。評判などの判別に使われる手がかりは、以降、**素性**と呼ぶ。アスペクトには、“カメラ-レンズ-ズーム機能”や、“携帯電話-サービス-保証”のような、**階層構造**があることが指摘されており [12], [13]、抽出後の構造化も課題となっている。

また、商品やサービスの意見は、ドメイン（たとえば、ホテルの評判、デジタルカメラの評判、書評など）が絞り込まれているため、肯定・否定の分類は、ニュースなどに比較して容易であるものの、各ドメインごとに文書中で使用される概念や語彙の傾向は異なることから、あらゆるドメインに対応した意見分析システムを構築するのは手間である。この問題に対応するために、ドメインに適応した意見分析システムについての研究がさかんに行われている。さらに、商品推薦のための協調フィルタリングサービスにおけるコールドスタート問題の解決や、偽装意見に代表されるスパム問題の解決などが、新たな課題として取り組まれている。

最後に、ブログやマイクロブログを対象とした意見分析では、さまざまな話題について、世論の傾向を明らかにするタスクが考えられる。したがって、ある話題に関連したブログあるいはマイクロブログについて、ある程度の規模のデータの集合を収集し、その話題に対する意見を分析する必要がある。また、マイクロブログにおいては、リアルタイムに情報が発信されるため、今現在起こっている出来事についての意見の分析を行える点が特徴となっている。書き方の特徴としては、顔文字などのインフォーマルな表現の使用に加えて、新しい特有の用語が次々と導入される点があげられる。また、ある話題についての意見も、時期によって変化する（意見ドリフト）。したがって、あらかじめ人手によるアノテーションが付いた意見分析コーパスは、構築しても必ずしも有意義とはいえず、また、*Mechanical Turk* のようなクラウドソーシングサービスを利用した場

合には、アノテーションの品質が問題となるとされている。

以上で見たとおり、意見分析の研究は、文書ジャンルに応じて要求されるタスクが異なり、また文書に現れる意見の性質も異なる。本稿では、各文書ジャンルの代表的な意見分析コーパスについて紹介し、現在、各文書ジャンルの意見分析コーパスにおいて必要とされている要件を明らかにする。続いて、それらの意見分析コーパスで定義された属性を活用した研究、文書ジャンルに応じたコーパスの活用方法、新たな要件を必要とする意見分析の研究にまつわるコーパスの構築事例について紹介する。これにより、文書ジャンルの性質を考慮したうえで、研究目的に応じてコーパスにアノテートすべき属性や、コーパスを活用することにより可能となる応用研究を明らかにする。

一方で、意見分析コーパスを専門家により作成するコストが高価であることから、クラウドソーシングサービスなどを利用し、コストをかけずにコーパスを作成する取り組み [14] や、HTML における手がかり語や顔文字など表層的な手がかりを利用して自動的に構築する取り組み [15], [16] が行われている。クラウドソーシングサービスを利用する方法では、複雑なインストラクションをとまなわない比較的単純な判定作業であることが望ましい。したがって、ニュースなどに現れる複雑な意見を対象とする場合には、先に説明したとおり、複数のアノテータによる判定のゆれについてチェック [9] し、他の判定者との一致率が低い判定者のアノテーションを除去する [14] などの工夫が必要となる。一方で、自動構築の場合には、獲得する表層パターンへの依存性による意見のタイプのカバレッジの制約^{*1}がある。また、判定誤りを含むようなアノテーションについては何らかのフィルタリングが必要となる。本サーベイでは、意見分析コーパスの構築の課題を解決するための取り組みについて紹介することで、コスト削減と品質の高いアノテーションを実現するための方針を提供する。

本稿の構成は以下のとおりである。2章では、意見分析コーパスの代表的な事例とその特徴を紹介する。3章では、意見分析についての最新の研究におけるコーパスの活用状況や、従来の枠組にはとどまらない新たなコーパスの要件を明らかにする。次に、4章で、意見分析コーパスの構築の課題に対する取り組みについて紹介をする。最後に、5章で、まとめと今後の展望について述べる。

2. 文書ジャンルに応じた意見分析コーパス

本章では、文書ジャンルごとに、最近の研究でも用いられることの多い利用可能な意見分析コーパスを紹介し、各文書ジャンルでコーパスに必要とされる基本的な要件を明らかにする。

2.1 ニュース文書を主な対象とした意見分析コーパス

本節では、ニュースを対象とした代表的な意見分析コーパスとして、MPQA 意見コーパスと NTCIR 多言語意見分析コーパスなどの、アノテーション方針に焦点を当てて紹介する。

(1) MPQA 意見コーパス

最初に、ニュース文書を主な対象とした代表的な意見分析コーパスとして、MPQA (*Multi-Perspective Question Answering*) 意見コーパス^{*2}を紹介する。本コーパスは、2002年に、Advanced Research and Development Activity (ARDA, 現, Disruptive Technology Office) の Northeast Regional Research Center (NRRC) が主催した夏のワークショップにおいて、従来の質問応答システムとは異なる多観点の質問応答を対象としたコーパスを開発したことに端を発する [18]。具体的なタスクとしては、政府機関で働く情報分析者が行う作業を自動化することを目的としており、ニュース記事から意見の断片を表すテキストを抽出し構造化することで、イラク戦争に対する適切な対応についての多様な意見を抽出したり、米国の京都議定書への対応に対して日本人が同意しているか、などの意見(見方)を問う質問に対する回答を提供したりできるような応用を検討していた。

この時点で、複数の研究グループに所属する10人の研究者の協力により、8つのトピック(のちに、10トピック)に関連する World News Connection (WNC)^{*3}の575の記事(Ver1.2で、535記事に選別)を対象としたコーパスが作成された。このコーパスは、Version 2.0で、692文書に拡張されており、その際には、Wall Street Journal, American National Corpus, Enron 社破たんに関する社員の電子メール、話し言葉の書き起こし、アラブ言語の翻訳などの文書が追加されている。この研究の貢献の1つは、意見情報のアノテーションのフレームワークを、多数の判定者による実験を通して厳密に定めた点にある [5]。アノテーションの方針としては、(1) 単語などを手がかりとして固定化した規則を用いない、(2) 文脈を考慮して判定する、(3) 方針を一貫させる、などがあり、サンプルを使ってアノテータを訓練することにより、アノテート方針を自分だけではなく他人とも一貫させるように訓練することにより、方針のずれを修正することを重視している。アノテータ間の一貫性の判定には、 κ 係数 [19] を用いている。これらの方針は、その他の意見分析コーパスも、構築の際に基本的には参考にしている。アノテーションに際しては、シェフィールド大学によ

^{*1} たとえば、顔文字を手がかりとして感情を判定する場合には、少数の感情を除いて対応関係があいまいであるという課題 [17] がある。

^{*2} <http://www.cs.pitt.edu/mpqa/>

^{*3} <http://wnc.fedworld.gov/>

る *GATE* システム^{*4}を使用している。意見情報については、以下の2つのタイプをもとに判定を行っている。なお、“心理状態 (private states)”とは、“ある経験をした人が、(ある対象に対して) ある態度を保持している状態”と定義される [5]。

(a) 心理状態の明示的な言及 (direct subjective)

以下に、例文を示す。

- 裁判官は、和解による早期解決を望んでいた。
- 財務省幹部は「今回の米国経済のマヒの影響は計り知れない」という。

このタイプの意見表現の、属性としては、以下のものが提案されている。

- (i) 意見のソース (意見を表明する人またはグループ。この場合は「裁判官」や「財務省幹部」などで、「著者」が意見を引用することで何かの態度を表明していると考えられる場合には、複数のソースが入れ子になってソースとして付与される)
- (ii) 意見の対象 (心理状態の対象または話題)
- (iii) 意見の強さ (弱い, 通常, 強い, 強烈)
- (iv) 意見の態度 (肯定, 否定, 両方, なし)
- (v) 発話の暗黙性
- (vi) 条件節などに含まれる意見の不明瞭さ

このうち、“意見の強さ”については、心理状態 (感情や信念の確信) の強さを手がかりとして判定しており、意見の態度が肯定または否定となる場合には、意見の強さは“通常”以上となる。上の1つ目の例文の場合には、例文の場合には、“通常”または“強い”となる (文脈に依存する)。“意見の態度”は、通常の極性 (ポジネガ) であるが、複数の極性が両立する場合と、極性がない場合を区別して付与する点に特徴がある。“発話の暗黙性”は、意見のソースがテキスト中に明示的に現れない場合 (具体的には著者となる場合) にフラグが立つ。上記の例文の場合には、意見表現に対しては、フラグが立たない。

(b) 主観に関連した手がかり表現の集合 (expressive subjective element)

以下に、例文を示す。

- 今後の情勢について専門家は、強権的なイメージが強く、国民に不評の新国王が、今後、どういう姿勢を打ち出すのか、見極めようとしている。

このタイプの意見表現の、属性としては、以下のものが提案されている。

- (i) 意見のソース (意見を表明する人またはグ

ループ。この場合、「専門家」に加えて、この意見を引用している「著者」が、「強権」「不評」などの言葉を引用することにより否定的な態度を示していると考えられる場合には、それぞれの要素もソースとして付与される)

- (ii) 意見の強さ (弱い, 通常, 強い, 強烈)

- (iii) 意見の態度 (肯定, 否定, それ以外, なし)

また、もう1つの貢献は、これらのアノテーションをもとに意見を集約するなどのアプリケーションを提案することであった [20]。その他、意見のソース (意見保有者) の判定 [21] や、文脈を考慮した極性判定 [22]、質問応答 [23] などの研究も進んでいる。最近では、提案手法を先行研究と比較する目的 [24], [25], [26], [27] や、あるいはラベル付きデータのリソース [28]、辞書データを利用した比較対象 [29] として利用されている。

一方、*MPQA* 意見コーパスのサイト^{*5}では、上記コーパス以外のデータとして、手がかり語辞書や意見分析システムが提供されている。また、*MPQA* 意見コーパスに *WordNet* の語義が付与されたもの [30] や、商品ディベート [31]、政治ディベート [32] へのアノテーションデータが提供されている。

(2) *NTCIR* 多言語意見分析コーパス

多言語を対象とした意見分析コーパスとして、*NTCIR* 多言語意見分析コーパス^{*6} [33], [34], [35] を紹介する。このコーパスは、日本語・英語・中国語の3カ国語 (ただし、中国語は繁体字と簡体字の2種類の言語を対象とする) のニュース記事を対象として、共通のトピックについて集めた記事を対象としたコーパスを作成している。コーパスのアノテーションの方針は、文献 [34] の付録に詳細が掲載されている。*MPQA* コーパスと比較すると、その構築方針 [5] に基本的には従っているが、仮説となるような意見を除くためのガイドラインが設定されていたり、意見保有者や意見対象のテキスト範囲が詳細に定義されていたりする点、極性は意見性がある文にだけ付与される点、1文に複数の意見を含む場合に分割して情報を付与する点などに特徴がある。また、これらの定義を3つの言語について共有している点も特徴となっている。本コーパスの応用目的は、多言語 (多地域) 間で共通の話題についての意見の差異を明らかにする点にある。タスクは3回行われており、その変遷を表 1 に示す。表中の+は新しい言語あるいはサブタスクが追加されたことを、-は変化がないことを示す。

ここでは、最新の *NTCIR-8* のサブタスクについて説明する。*NTCIR-8* では、以下に示す5つの要素技術サブタスクと、1つの応用サブタスクが行われた。

*4 <http://gate.ac.uk/>

*5 <http://www.cs.pitt.edu/mpqa/>

*6 <http://research.nii.ac.jp/ntcir/data/data-ja.html>

表 1 NTCIR-6, 7, 8 における多言語意見分析タスクの変遷

Table 1 Evolution in NTCIR-6, 7, 8 multilingual opinion analysis task.

| NTCIR | 言語 | サブタスク | 判定単位 | 想定アプリケーション | コーパス | 新聞の期間 |
|---------|-----------------|-----------------------------------|------|------------|--|-----------|
| NTCIR-6 | 英語, 日本語, 繁体字中国語 | 意見文判定 適合文判定 極性判定 意見保有者判別 | 文 | 情報検索 | 毎日新聞, 読売新聞 Xinhua English, Hong Kong Standar, CIRB など | 1998-2001 |
| NTCIR-7 | +簡体字中国語 | +意見対象判別 | 意見節 | 質問応答 | +Xinhua Chinese | - |
| NTCIR-8 | - | +言語横断意見質問応答 | - | 意見質問応答 | +NYT, UDN | 2002-2005 |

表 2 NTCIR-8 多言語意見分析タスクのテストコレクションのサイズ

Table 2 Test collection size in NTCIR-8 multilingual opinion analysis task.

| 言語 | トピック数 (意見質問数) | | | 文書数 | | | 文数 | | | 意見節数 | | |
|--------|---------------|------|--------|-----|------|--------|-------|------|--------|-------|------|--------|
| | 合計 | サンプル | テストデータ | 合計 | サンプル | テストデータ | 合計 | サンプル | テストデータ | 合計 | サンプル | テストデータ |
| 日本語 | 21 | 1 | 20 | 178 | 8 | 170 | 6,992 | 322 | 6,670 | 7,894 | 341 | 7,553 |
| 英語 | 21 | 1 | 20 | 150 | 12 | 138 | 6,564 | 399 | 6,165 | 6,723 | 500 | 6,223 |
| 簡体字中国語 | 20 | 1 | 19 | 410 | 25 | 385 | 4,720 | 228 | 4,492 | 4,744 | 232 | 4,512 |
| 繁体字中国語 | 21 | 1 | 20 | 787 | 12 | 775 | 9,684 | 160 | 9,524 | N/A | N/A | N/A |

(a) 意見文判定サブタスク

MPQA 意見コーパスの構築方針 [5] に基づき人手で判定を行ったものを正解として、自動抽出する。

(b) 適合意見文判定サブタスク

ある話題に適合する意見文について、自動抽出する。

(c) 極性分類サブタスク

1つの意見文中の極性が異なる節を分割することで、意見節を定義し、その節を、肯定意見、否定意見、中立意見のいずれかに分類する。

(d) 意見保有者判別サブタスク

意見を表明する人・組織・団体を判別する。

(e) 意見対象判別サブタスク

意見の対象・トピックを判別する。

(f) 言語横断意見質問応答サブタスク

英語の質問に対して、回答意見を異なる言語の文書集合から検索する。

これらのサブタスクの遂行のため、NTCIR-8 のサブタスクでは、意見を問う 20 の質問を用意し、毎日新聞 (日本語), *New York Times* (英語), *United Daily News* (繁体字中国語), *Xinhua Chinese* (簡体字中国語) の文書から回答となる意見を集約するため、サブタスクに関係する意見性, 意見適合性, 極性, 意見保有者, 意見対象, 回答性をアノテートしたコーパスを作成した。コーパスのサイズを表 2 に示す。

最近では、これらのデータは、提案手法の有効性の評価 [36] や、複数言語のリソースを活用した意見分析 [28], 言語を横断した意見分析 [37] などの研究に用いられている。

(3) 上記 2 つ以外で、ニュース文書を対象とした意見分

析コーパスとしては、*SemEval-2007 Task 14: Affective Text* [38]^{*7} や、*Sentiment-annotated quotation set* [39], [40]^{*8} がある。前者は、1,250 のニュースサイトおよび新聞記事から収集した見出しについて、6 つの感情ラベルと 2 つの極性ラベルを付与している。後者は、*European Commission* の *Joint Research Centre* が開発した *Europe Media Monitor (EMM)* [41]^{*9} でクロールしたニュース記事のうち、意見の保有者と対象が明確な 1,590 の引用意見 (quotation) について、意見情報をアノテートした結果を公開している。

2.2 レビュー文書を対象とした意見分析コーパス

レビューを対象とした意見分析コーパスは、ニュースと比較すると、詳細なアノテーションを必要としない。具体的には、意見保有者は、多くの場合、著者と判定される。また、ニュースのようにあいまいな言い方が少なく、意見の対象と書き込む目的が明確である影響から、肯定・否定の表現もはっきりしている。さらに、文書単位の肯定・否定情報や、商品の評価 (5 つ星) は、レビューサイトの評判を再利用することで、ある程度対応できる。以上から、アノテータによるアノテーションが必要な情報は、(1) 商品評価の аспекト、(2) 文単位の極性、(3) 偽装意見などの特殊な評価、(4) 意見分析に必要な手がかり、などが対象となる。また、情報要求が各レビューサイトごとに少しずつ異なることから、研究者は自分の要求に合ったコーパスを自分で構築することが多い。一方で、以下で紹介するコーパスは、極性の判別やアスペクト表現の抽出などの要

*7 <http://www.cs.unt.edu/~rada/affectivetext>

*8 http://langtech.jrc.ec.europa.eu/JRC_Resources.html

*9 <http://emm.newsbrief.eu/overview.html>

(肯定) if you sometimes like to go to the movies to have fun , wasabi is a good place to start .
 (否定) simplistic , silly and tedious .

図 1 映画レビューを対象とした意見分析データセット [44] の例
Fig. 1 Examples in movie sentiment dataset [44].

picture[+2]##they fired away and the picture turned out quite nicely . (as all of my pictures have thusfar) .

図 2 Amazon レビューを対象とした製品の属性に基づく要約データセット [46] の例
Fig. 2 Example in feature-based summarization dataset [46].

素技術に関する新たな提案について、先行研究との精度の比較評価に用いられることが多い。本節では、レビュー文書を対象とした4つのコーパスを中心に紹介する。

(1) 映画レビューを対象とした意見分析データセット

まず、映画レビューの意見分析データセット^{*10}を紹介する。本コーパスは、*Internet Movie Database (IMDb)*^{*11}を対象とした意見分析コーパスである [42]。具体的には、肯定・否定それぞれ1,000のレビュー文書 [43] や、肯定・否定それぞれ5,331のレビュー文 [44] を提供している。その他、レビューにおける極性の段階的評価 (意見の強さ) [44] や、5,000 ずつの主観文・客観文 [43] のデータセットも提供している。肯定・否定文のアノテーションの具体例を図 1 に示す。最近では、提案手法を先行研究と比較する目的で、利用されている [26], [45]。

(2) Amazon レビューを対象とした製品の属性に基づく要約データセット

次に、Amazon レビューを対象としたデータセットとして、製品の属性に基づく要約 (Feature-based Summarization)^{*12}を紹介する。本コーパスは、最初に5つの製品 [46]、のちに9つの製品を追加して [47]、Amazon のレビュー文に、その文に現れる製品の属性と、それに対する評価値を + (肯定) / - (否定) のかたちで付与している。その他、意見スパム (偽装レビュー) [48] や、製品を比較する意見文 [49], [50] のアノテーションデータも公開している。属性と評価値のアノテーションの具体例を図 2 に示す。この例は、デジタルカメラ (Canon G3) を対象としており、“picture” が属性、“+2” が評価値、“##” の後がレビューを示している。最近では、アプリケーションのリソース [51], [52] として用いられている。

(3) 複数のドメインを対象とした意見分析データセット

^{*10} <http://www.cs.cornell.edu/people/pabo/movie-review-data>
^{*11} 当初は、ニュースグループ rec.arts.movies.reviews のアーカイブ、最近は、Amazon.com により運営されている。
<http://www.imdb.com/>
^{*12} <http://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html>

```
<review>
<product_name>Why We Fight: DVD: John
McCain,...</product_name>
<product_type>dvd</product_type>
<helpful>6 of 9</helpful>
<rating>5.0</rating>
<reviewer>Mr. James A. Blair "UK
Geek"</reviewer>
<review_text>I recomend it, not so much as some
dark conspiracy, but more a genuine question...
</review_text>
</review>
```

図 3 複数のドメインを対象とした意見分析データセット [53] の例 (抜粋)
Fig. 3 Example in multi-domain sentiment dataset [53].

また、Amazon レビューの複数のドメインを対象としたデータセット^{*13}を紹介する [53]。このデータセットでは、書籍、DVD、電気製品、台所用品の4つのドメインに関する製品データを取り扱っている。それぞれのドメインでは、1,000の肯定・否定のラベルが付いたレビュー文書と、ラベルの付いていない、より多くのレビュー文書が含まれている。提供されているレビューの具体例を図 3 に示す。

最近では、ドメイン適応の検証 [54], [55], [56], [57] や、提案手法の有効性の評価 [58]、バランス不均衡なデータの分類 [59]、レビューの品質の判定 [60] に用いられている。

(4) 楽天データセット

楽天技術研究所では、楽天市場の約5,000万件の商品についての約1,660万件のレビュー、楽天トラベルの11,468件のホテル・旅館についての約34万件のレビュー、楽天GORAの1,669件のゴルフ場の約32万件のレビューを、研究利用目的のデータセットとして、国立情報学研究所^{*14}および高度言語情報融合フォーラム (ALAGIN)^{*15}から公開している。また、そのデータを利用して、2008年^{*16}、2009年^{*17}、2010年^{*18}、2011年^{*19}と4回にわたり、研究開発シンポジウムを開催している。

(5) その他、レビューに関する数多くのコーパス (たとえば、SFU レビューコーパス [61] や、TripAdvisor Dataset [58], [62]^{*20}など) が存在するが、上記の4つ

^{*13} <http://www.cs.jhu.edu/~mdredze/datasets/sentiment/>
^{*14} <http://www.nii.ac.jp/cscenter/idr/rakuten/rakuten.html>
^{*15} <http://alaginrc.nict.go.jp/>
^{*16} <http://rit.rakuten.co.jp/conf/rrds2008/>
^{*17} <http://rit.rakuten.co.jp/conf/rrds2009/>
^{*18} <http://rit.rakuten.co.jp/conf/rrds3/>
^{*19} <http://rit.rakuten.co.jp/conf/rrds4/>
^{*20} <http://times.cs.uiuc.edu/~wang296/Data/>

表 3 TREC 2006/2008 ブログテストコレクションのサイズ
Table 3 TREC 2006/2008 blog test collection size.

| サイズ | Blogs06 | Blogs08 |
|---------------|-----------------------|-----------------------|
| ブログの数 (重複は排除) | 100,649 | 1,303,520 |
| クロールの期間 | 2005年12月6日-2006年2月21日 | 2008年1月14日-2009年2月10日 |
| パーマネントリンクの数 | 3,215,171 | 28,488,766 |

と比較すると、利用規模が小さいことから、ここでは紹介を割愛する。

2.3 ブログを対象とした意見分析コーパス

ブログを対象とした意見分析コーパスは、CGMを対象とした応用に共通する性質として、多種多様な大量のデータを含むコーパスへの期待が高く、これまで述べてきたコーパスとは、少し性質が異なる。本節では、このような性質を満たすコーパスとして、TREC ブログコーパスと関連研究について紹介する。次に、ブログを用いた意見分析の応用のために設計されたTAC 意見質問応答・要約タスクについて紹介する。最後に、マイクロブログを対象としたならびにその他のコーパスについて紹介する。

(1) TREC 2006/2008 ブログコーパス

情報検索に関する国際評議会 TREC (Text REtrieval Conference) において、2006年から2010年にわたり、ブログトラック [63] が開催された。このためのテストコレクションが、グラスゴー大学より配布されている*21。テストコレクションは、2006年と2008年に作成したものの2種類があり、その内容を表 3 に示す。このデータを用いて、意見発見 (opinion finding) タスク、ブログ抽出タスク、トップニュース検出タスクなどが行われている。意見発見タスクでは、ある対象 (人名、地名、製品名など) に対する意見を表明するブログポストの検索技術の評価しており、従来の情報検索に基づく適合性 (relevance) と、意見性 (opinionatedness) に基づくランキングを評価している。本タスクでは、3年間の評価により、150のトピックに適合する175,818のブログポストに対して、人手評価が提供されている。本タスクのデータを用いて、最近では、ブログフィード (同一サイトに属するブログポストの集合) についての意見性判定モデル [64] の提案が行われている。

(2) TAC 2008 意見質問応答・要約タスク

TAC (Text Analysis Conference) では、2008年度に、意見質問応答タスクならびに意見要約タスクが開催された [65] *22。このタスクでは、TREC Blog Track で使用されたトピックを参照しつつ、ある対象 (製品名、

- (a) What motivated negative opinions regarding purchasing a car from CARMAX? What motivated positive opinions of CARMAX from car buyers?
- (b) What reasons are given for liking the services provided by Jiffy Lube? What reasons are given for not liking the services provided by Jiffy Lube?
- (c) Why do people like Starbucks better than Dunkin Donuts? Why do people like Dunkin Donuts better than Starbucks?

図 4 TAC 2008 [65] において使用された質問の例
Fig. 4 Question examples in TAC 2008 [65].

人名、組織名など) について書かれた意見を問う質問に対して、その回答を TREC 2006 ブログテストコレクションから抽出し (意見質問応答タスク)、回答を集約して流暢なテキストを生成する技術 (意見要約タスク) を評価している。意見質問応答タスクでは、50の対象に対して、90の厳密な回答を持つリスト型質問 (rigid list question) と87の緩い回答を持つリスト型質問 (squishy list question) に対する回答抽出技術の評価している。前者は文字列の完全一致を、後者はナゲット [66] と呼ばれる情報単位を回答の単位として定義し、概念的な一致をもとに評価を行う。意見要約タスクでは、意見質問応答タスクのうち、22の対象についての緩い回答を持つリスト型質問を選択し、要約を生成する技術を、内容、読みやすさ、回答性の観点から評価している。TAC 2008 において使用された質問の例を図 4 に示す。

(3) TREC 2011 マイクロブログコーパス

TREC では、2011年度よりマイクロブログに焦点を当てたトラック*23を開始した。このためのテストコレクションは、NIST*24から配布されている。このデータを用いて、リアルタイム検索に焦点を当てたタスクが行われている。

(4) その他のコーパス

EmotiBlog [67], [68] は、英語、スペイン語、イタリア語のブログを対象とした多言語意見分析コーパスで、京都議定書、ジンバブエのムガベ大統領、2008年の米国大統領選挙などのトピックを対象としている。本コーパスは、文献 [67] の著者へのリクエストにより利

*21 http://ir.dcs.gla.ac.uk/test_collections/

*22 <http://www.nist.gov/tac/data/past/2008/OpSummQA08.html>

*23 <http://sites.google.com/site/microblogtrack/home>

*24 <http://trec.nist.gov/data/tweets/>

用可能である。中国では、*Chinese Opinion Analysis Evaluation (COAE)* [69]^{*25}という会議が、2008-2011年にわたり3回開催されており、中国語のブログを対象としたコーパスを利用している。

日本語については、京都大学とNTTが協力して、4つのトピックを対象とした解析済みブログコーパス^{*26}を公開している。また、京都観光ブログデータ (1,041記事)が、高度言語情報融合フォーラム (*ALAGIN*)^{*27}から公開されている。また、最近の研究では、Ptaszynskiらにより構築された50億語のコーパス [16]があり、今後の展開が期待される。マイクロブログを対象としたコーパスでは、Goら [70]^{*28}により構築された極性分類のためのコーパス^{*29}が、引用されることが多い。このコーパスでは、訓練データは、顔文字を手がかりとしているためノイズ混じりのデータであるのに対して、テストデータは、顔文字とは無関係に、人手で極性を付与している。

3. 最近のコーパスの活用事例と新たな要件

本章では、文書ジャンルと研究目的に応じてアノテートする必要のある属性や、コーパスを活用することにより可能となる応用研究を明らかにするために、最近の研究を対象として、2章で紹介したコーパスにあるような基本的な要件に対する取り組み、コーパスの活用事例、2章のコーパスではカバーしきれない課題に対する、新たなコーパス構築の取り組みなどについて紹介する。

意見分析の研究は、2002年ごろから活発に行われており [18], [42], [71]、初期の研究は、レビュー文書を用いた教師あり学習 [42] や教師なし学習 [71] に基づく文書単位の肯定・否定の分類が活発に行われていた。一方で、ニュース文書を対象とした意見分析コーパスの作成も同時期に始まっている [18]。その後、数多くの国際ワークショップ (*AAAI 2004 Spring Symposium: Exploring Attitude and Affect in Text: Theories and Applications* [72]; *COLING-ACL 2006 Workshop: Sentiment and Subjectivity in Text; Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA) 2010-2012; Sentiment Analysis Symposium*^{*30}など) や評価型ワークショップ (*TREC Blog Track 2006-2010; NTCIR Multilingual Opinion Analysis Tasks; SemEval-2007: Affective Text; TAC 2008 Opinion Question Answering and Summarization Tasks; TREC Microblog Track 2011*-など) を

経て、発展を続けている。また、日本でも、2005年の第12回言語処理学会年次大会併設ワークショップ「感情・評価・態度と言語」、2007年の電子情報通信学会言語理解とコミュニケーション研究会主催の「主観表現処理の最前線」シンポジウム、2008年以降の楽天株式会社主催の「楽天研究開発シンポジウム」など、着実に研究の場を広げている。これらの研究の変遷は、文献 [1], [2], [3] などにまとめられている。

意見分析は、2007年ごろまでは、辞書あるいはシード語を活用した分類手法 [71] と、コーパスを用いた教師あり学習による分類手法 [42] が、主流であった。しかし、ラベル付き文書をドメイン別に人手で準備する手間が大きい点を考慮し、ラベルなし文書あるいは辞書と機械学習を組み合わせた半教師あり学習の方法に関する提案が増えてきている。これらの状況をふまえ、意見分析についての最近の研究におけるコーパスの要件と活用事例を明らかにするために、自然言語処理、情報検索の国際会議 (ACL 2011-2012, EMNLP 2011-2012, SIGIR 2011-2012, CIKM 2011-2012, WSDM 2011-2012, EACL 2012, WWW 2011-2012) で2011年以降に発表された50件の意見分析の研究について、文書ジャンルごとに、研究内容と使用コーパスを整理した結果を以下に示す^{*31}。

3.1 ニュースを対象とした研究

ニュースを対象とした研究については、1章でも述べた文書ジャンルの特徴から、以下のように整理して説明する。

(1) 意見表現 [5] の定式化に基づく意見分析の研究

1章でも述べたとおり、ニュース記事からの意見を分析するために、さまざまな識者や組織により表明されている意見を話題別に集約するタスク [8] が提案されている。この応用に向けて、従来の意見分析コーパス [5] でも対象とされてきた、「誰が何についてのどういった態度の意見を表明した」という意見表現 [5] の定式化に基づく、極性とその強さの判別、意見表現、意見保有者の抽出といった要素技術の研究が行われている。

極性については、Zhouら [36] と Yessenalinaら [24] による取り組みがある。Zhouらは、「フジモリ大統領は、国際社会からは非難されたが、腐敗した支配層を嫌っていた国民からは、敬愛されていた。」というような文章において、手がかり語 (「非難」「腐敗した」「嫌う」) の多さから否定的と分類されてしまうところを、*RST (Rhetorical Structure Theory)* により文節間の関係を考慮して肯定的と分類することで、極性分類の精度が向上することを明らかにした。極性のアノテーションは、2.1節で紹介した、*NTCIR* 多言語意見分析

^{*25} <http://www.ir-china.org.cn/Information.html>

^{*26} <http://nlp.ist.i.kyoto-u.ac.jp/kuntt/>

^{*27} <http://alaginrc.nict.go.jp/>

^{*28} <http://www.sentiment140.com/>

^{*29} <http://cs.stanford.edu/people/alecmgo/trainingandtestdata.zip>

^{*30} <http://sentimentsymposium.com/>

^{*31} 2012年11月までの、自然言語処理・情報検索のトップコンファレンスに限定している。ショートペーパーは除外している。

コーパス^{*32}を活用し、さらに、RSTに基づく談話関係のアノテーションを自ら追加した。Yessenalinaらは、“very good”（非常によい），“not perfect”（完全ではないが）など、形容詞と周辺語を組み合わせて句を形成する際に、意見の強さがどのように変化するかを、行列空間モデルで定式化し、極性の強弱を分類した。極性の強さは、2.1節で紹介した、MPQAコーパスのアノテーションを用いている。

意見保有者については、構文的にさまざまな形態をとるため、意見分析コーパス中にすべてのパターンが訓練データとして含んでいるとはいえないケースがある。Wiegandら[25]は、ニュースにおける意見保有者は、格役割でいう、被影響者（patient）の位置に現れる意見保有者が、動作主（agent）の位置に比べて少ないことに着目し、辞書を用いた汎化手法を提案した。実験には、MPQA意見コーパス^{*33}を訓練データとして採用し、小説を対象として意見保有者を付与した意見分析コーパスをテストデータとして構築した。意見表現については、CRFに基づく抽出[73]において、現在と直前のトークンを手がかりとしてラベル付けを行っているのに対して、Yangら[27]は、構成素や句といった構文単位を手がかりとするように拡張した。意見表現は、MPQAコーパスのアノテーションを用いている。

(2) 多言語コーパスに関する研究

意見分析コーパスは、さまざまな言語について構築可能である。ここでは、複数の言語のコーパスを対象として、極性を判定する研究を紹介する。Luら[28]は、言語間で対応する文は、その極性や強さも類似しているという前提に立ち、複数の言語の意見分析コーパスを活用して、それぞれの極性の判別の精度を向上させた。ラベルつきコーパスとしては、MPQA意見コーパスとNTCIR多言語意見分析コーパスを訓練データとして使用し、ラベルなしコーパスとしては、中国語と英語のニュース対訳コーパスを用いて、極性を潜在変数として、両言語から推定する極性の同時確率を最大化することにより、極性を判別した。Mengら[37]は、Luら[28]と同一のデータを用いて、言語を横断した混合生成モデルを提案し、ラベルなしデータから極性の手がかり語を広く学習することにより、対象言語の極性を判別した。

(3) 新たな意見分析の定式化

1章でも述べたとおり、ニュース中の意見は、肯定や否定に必ずしも分類できないという立場から、新たな意見分析手法が提案されている。

Parkら[74]は、論争が起きているようなテーマ、たと

えば、政党Aが政党Bを批判する一方で、政党Bがそれは誤解に基づいていると主張した場合について、双方の主張を、肯定や否定あるいは賛否という形での分類をするのではなく、それぞれの立場を単位として記事を分類した。評価データは、韓国語のニュースを対象として、アノテータ3名が異なる立場を抽出し、極性の分類よりも安定して判別できることを示した。Fangら[75]は、立場の異なる記事集合から、LDA[76]に基づき、共通のトピックについての、代表的な意見語を抽出し、立場ごとの差異を明らかにした。評価データは、民主党・共和党の上院議員の発言ならびに、米国・インド・中国の新聞を対象としており、意見語についてのアノテーションは行っていない。

3.2 レビュー文書を対象とした研究

レビューを対象とした意見分析は、商品やサービスについての大量のレビューにおける評判を、顧客または商品の開発業者が確認をするために行われる。このため、レビュー文書あるいは、レビューに出現する商品の細かい特徴（1章で説明したアスペクト）の極性（肯定・否定）やその強さ（5つ星判定）を判別する研究が最もさかんであり、2.2節で紹介した意見分析コーパス[42]、[43]、[46]もこうした情報を備えている。ここでは、こうした従来のコーパスにも存在する属性を判別する研究と、意見分析コーパスの活用する研究、従来のコーパスには存在しない新たな観点を提供する研究、などに分類して詳細を説明する。

(1) 極性の判別やアスペクトの抽出に関する研究

レビュー文書の極性の判別は、長らく研究が進められており、新たな視点としては、従来の研究ではとらえきれない素性を定義することと、教師なし学習により取り組む点が研究課題となっている。Balamuraliら[77]は、WordNetの語義を素性とするすることで、極性を判別する手法を提案した。また、訓練データに、テストデータで極性を判別するための手がかりがない場合に、語義の類似度を考慮することで、精度の向上につながることを明らかにしている。実験には、Yahooの旅行ドメインレビューを用いている。Maasら[45]は、ラベルつきコーパスにおける単語の出現から、極性に関して類似度を計算することのできる単語のベクトル表現を学習し、極性を判別した。実験には、2.2節で紹介した、Pangらによる肯定・否定レビューのデータ[43]を拡張し、より大規模な映画レビューコーパス^{*34}を作成した。極性は、レビューの文書を単位として、10点満点のうちの4点以下を否定的（ネガティブ）、7点以上を肯定的（ポジティブ）なレビューとして定義し、5-6点のレビューは除外している。Bespalov

^{*32} <http://research.nii.ac.jp/ntcir/data/data-ja.html>

^{*33} <http://www.cs.pitt.edu/mpqa/>

^{*34} <http://www.andrew-maas.net/data/sentiment>

ら [58] は、通常の 1-グラムや 2-グラムより優れた素性として、ニューラルネットワークを利用して、 n -グラム表現を潜在的な素性として学習することにより、意見の極性と強さを判別する手法を提案した。コーパスは、Amazon や、TripAdvisor を対象とした大規模データセット^{*35}を用いた。TripAdvisor における極性は、部屋、位置、清潔さなどさまざまなアスペクトについての評価もあるが、欠落も多いことから、文書全体を対象として評価しているが、より細かい単位の意見表現を対象として極性を判別する場合には、アノテータによる極性と強さの付与が必要となる。

一方、アスペクト表現の抽出は、レビュー文書を対象とした意見分析の中で、現在最もさかんな研究の一つである。Hai ら [78] は、アスペクト表現と意見表現、あるいはアスペクト表現どうしの共起関係を利用し、頻出アイテム集合マイニング手法とブートストラップを組み合わせた抽出手法を提案し、シード語としては、ドメインを表す 1 つのシード語（たとえば、“携帯電話” など）によりアスペクト表現が抽出できることを明らかにした。評価コーパスは、中国語の携帯電話とホテルのレビューを対象として、2 人のアノテータが意見の属性を付与して作成した。Liu ら [79] は、意見表現を手がかりとして、アスペクト表現と間のアラインメントを行うことにより、アスペクト表現を抽出した。これにより、共起関係を計算する際の単語間の距離の制約や、構文関係を考慮する際の構文解析の精度の問題を避けることができる。評価コーパスは、中国語のレビューデータセット (2.3 節で紹介した COAE2008^{*36}) とレストランレビューを対象として 3 人のアノテータがアスペクト表現の付与を行った。また、2.2 節で紹介した英語のコーパス^{*37}も活用されている。

さらに、意見表現とアスペクト表現を組み合わせた抽出手法に関する研究 [29], [80], [81] が行われている。Moghaddam ら [80] は、LDA [76] に基づき意見表現とアスペクト表現を抽出する手法として、(1) 潜在変数を 1 つだけ設定するか、あるいはアスペクト表現と意見表現について別々に設定するか、(2) 文書中の単語すべてを対象としてモデル化するか、あるいは意見表現だけを対象としてモデル化するか、(3) アスペクト表現と意見表現とを独立関係とするか、あるいは依存関係とすることといった観点について、コーパスサイズも考慮しながら、どのような組合せが有効か、検証した。コーパスを用いた評価については、人手でアノテートした Amazon を対象とした製品の属性に基づく要約

データセット^{*38}を用いた精度評価と、Epinions.com から収集した 10 万件以上の大規模な商品レビューデータ^{*39}のパープレキシティによる評価との相関を調査し、自動収集データのパープレキシティ評価の妥当性を示した。Jo ら [81] は、(1) アスペクト表現と意見表現について、それぞれ潜在変数を仮定しており、(2) すべての単語を使用して、(3) 依存関係を仮定した LDA の拡張モデルを提案した。評価コーパスは、Amazon と Yelp^{*40}における製品とレストランレビュー^{*41}を対象としている。Lu ら [29] は、極性辞書と、ドメインに依存したレビューコーパスについて、同義語と反対語の関係や、接続関係を表す手がかり語などの制約を最適化することにより、アスペクトと極性の組合せを単位とした、意見表現の辞書を構築した。評価データは、TripAdvisor^{*42}から収集したホテルレビューとカスタマーレビューについて、2 人のアノテータにより、アスペクト（位置、食事、部屋、サービスなど）に依存した極性を付与することにより構築した。

また、同一のアスペクトは、複数のレビューにおいて、さまざまな表現（たとえば、“picture” と “photo” など）がある。Zhai ら [82] は、同一単語の共有、WordNet による単語間類似度を制約として、EM アルゴリズムに基づき、同義となるアスペクト表現を分類した。評価データは、意見分析をサービスとして提供する会社から、5 つのドメインレビューを収集し、顧客に同義のアスペクト表現を分類させた。Sauper ら [83] は、レビューのスニペットを対象とし、アスペクトとアスペクトについての極性を潜在変数として、スニペットを生成するモデルを考慮することにより、スニペットをアスペクトと極性ごとに分類した。評価コーパスは、Yelp 上の 1 つのレストランに対するレビューのスニペットの集合を対象として、共通のアスペクト表現に言及しているスニペットを、Mechanical Turk を使ってクラスタに分類することにより構築した。一方、Yu ら [13], [84] は、商品販売サイト上で項目とされているアスペクトの一覧とその関係を抽出した後、アスペクト表現間の、レビュー文書上の文脈類似度や共起関係などを考慮し、階層的に構造化している。評価コーパスは、cnet.com や viewpoints.com など、いくつかのレビューサイトから、11 の商品についてレビューを収集し、5 人のアノテータにより、アスペクト表現を抽出し、その階層構造を付与している。さらに、Yu ら [84] は、Yahoo! Answers から 220 の商品についての質問を収集し、アスペクトの階層構造を考慮して、下位の

^{*35} <http://mst.cs.drexel.edu/datasets/CIKM2011>

^{*36} <http://www.ir-china.org.cn/Information.html>

^{*37} <http://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html>

^{*38} <http://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html>

^{*39} <http://www.sfu.ca/~sam39/Datasets/EpinionsReviews/>

^{*40} <http://www.yelp.com>

^{*41} <http://uilab.kaist.ac.kr/research/WSDM2011>

^{*42} <http://www.tripadvisor.com>

アスペクトや明示的に記述されていないアスペクトについての意見も含んで回答する質問応答システムを実現した。回答は、3人のアノテータにより作成した。

(2) ドメイン横断型コーパスの研究

1章でも述べたとおり、レビューの意見情報は、ドメイン（たとえば、ホテルの評判、デジタルカメラの評判、書評など）が絞り込まれている。各ドメインごとに文書中で使用される概念や語彙の傾向は異なることから、あらゆるドメインに対応した意見分析システムを構築するのは手間である。ここでは、複数のドメインの意見分析コーパス^{*43} [53] を用いて、あるドメインにおけるラベルつき文書により学習した情報を用いて、別のドメインにおけるラベルなし文書を対象として意見分析を実現する研究 [54], [55], [56], [57] について紹介する。

Ponomareva ら [57] は、ラベルつき文書とラベルなし文書を対象として、極性についての文書間類似度に基づき結合されている文書間で、同一の極性を維持するようにラベルなし文書の極性を更新するグラフアルゴリズムを用いて、ラベルなし文書の極性を判別した。また、ドメイン間類似度とドメインの複雑さが増すほど、ラベルつき文書間の類似度のラベルなし文書間の類似度に対する重みパラメータを、大きめに調整する必要性を明らかにした。一方、He ら [54] は、パラメータ調整を必要としない手法を提案した。まず、トピックと極性を考慮して単語を生成する *JST* モデル [85] を、ラベルつき文書とラベルなし文書をあわせた文書群に適用し、各文書中の単語に割り当てられるトピックと極性の組合せを推定する。次に、ラベルつき文書について、文書中の単語から構成された素性空間を、単語に割り当てられたトピックと極性ラベルの組合せを素性として追加することにより拡張し、情報利得により有用な素性を選択した後、教師あり学習の訓練に用いる。ラベルなし文書は、出現する単語以外に、単語に対応するトピックと極性ラベルの組合せも素性として、その極性を推定する。また、ラベルなし文書の極性を *JST* モデルにより推定することにより、推定の信頼度の高い文書を、訓練データとして追加することもできる。これとは別に、Bollegala ら [55] も、ラベルなし文書を対象とした素性空間の拡張手法を提案している。まず、レビュー文書における素性（単語 1-グラムや 2-グラムなど）の共起情報と、レビュー文書の極性を考慮して、複数のドメイン間で極性を共有する単語群を登録したシソーラスを構築する。次に、別のドメインのラベルなし文書について、シソーラス中で極性を共有する単語群により素性空間を拡張し、その

極性を推定する。

最後に、Li ら [56] は、ドメインごとに、意見表現とアスペクト表現を抽出する手法を提案した。まず、一般的な意見表現 (“good”, “bad” など) とアスペクト表現とがアノテートされたドメインの文書から、その構文関係を抽出する。次に、別のドメインにおいて、一般的な意見表現と同様の構文関係を持つアスペクト表現を、シード語として抽出する。さらに、構文関係を考慮しながら、ブートストラッピングにより、意見表現とアスペクト表現を拡張する。評価コーパスは、Blitzer らのデータセット^{*44} [53] と、先行研究 [86] において用いた 500 の映画レビュー、601 の商品レビューを対象として、極性表現、アスペクト表現をタグ付けしたものをを用いた。

(3) 新たな意見分析の定式化

ここでは、2.2 節で紹介したコーパスでは研究に不十分な、新たな意見分析の応用として、意見スパムの検出 [52], [87], 協調フィルタリング [88], [89], 有用性 [60], 要約 [51], [90], 構造の抽出 [91] について紹介する。

レビュー文書の評判を参考する購入の判断をするユーザが増加するに従い、商品やサービスの評判を向上（あるいは低下）させるために、本来と異なるレビューを書くことが行われ、問題となっている。Ott ら [87] は、偽装レビューコーパスとして、*Mechanical Turk* を使用して作成させた 400 の偽装レビューと TripAdvisor から収集した 400 の真のレビュー^{*45} を準備し、*n*-グラムと心理言語学に基づく手がかりを素性とした単純なテキスト分類手法により、人間が判定するよりも偽装意見スパムを精度良く判定できることを明らかにした。これに対し、Mukherjee ら [52] は、協力して偽装レビューを執筆するグループに着目した。まず、Amazon レビューを対象として、頻出アイテム集合マイニングにより、同じ商品に同じ評価をするレビューアを偽装レビューア候補として抽出し、次に、グループ、メンバ、商品の関係に着目して偽装レビューアグループをランク付けした。評価コーパスは、8 人の専門家により、偽装レビューアグループを手動で判定させ構築した。

また、レビューアによる評価を推定することで、協調フィルタリングによる推薦へと応用できる。Kawamae [88] は、レビューアの嗜好、意見、アスペクトを潜在変数とした階層トピックモデルにより、レビューアによる今後のトピックに対するレビューを予測するモデルを提案した。評価コーパスは、*MSN*^{*46} から取

^{*43} <http://www.cs.jhu.edu/~mdredze/datasets/sentiment/>

^{*44} <http://www.cs.jhu.edu/~mdredze/datasets/sentiment/>

^{*45} <http://www.cs.cornell.edu/~myleott/op-spam>

^{*46} <http://movies.msn.com>

集した映画レビューとレビューのデータを使用し、レビューの履歴から、推薦のカバレッジなどを評価した。また、Moshfeghiら[89]は、コールドスタート問題に代表される評価データが不十分な状況における映画の推薦のために、あるユーザが映画を好きな確率を推定した。確率の推定には、映画が好きなユーザのグループを潜在変数としたLDAを採用し、訓練データは、IMDb^{*47}から映画レビューを収集し、素性は、映画のレビューから抽出した感情と、映画の登場人物やジャンルを用いた。推薦の評価コーパスは、MovieLensデータセット[92]を用いた。

レビュー文書を評価する別の観点として、そのレビューが購入に役立ったか否かという有用性 (Helpfulness) がある。Hongら[60]は、人手で有用性が判別されていないレビューの有用性を推定するために、アスペクト表現の出現による情報要求の充足、意思を表す助動詞や、過去形などによる経験表現、商品の極性の強さの平均値からの差分を素性として、Ranking SVMを用いて有用性を自動推定した。評価コーパスは、Blitzerらのデータセット^{*48}[53]に出現する商品に関するレビューを、Amazonから収集した。有用であるか否かは、Amazonユーザの「役に立った」の投票率が6割を超えるか否かで分類してアノテートし、有用性によるレビューのランキングは、著者が判定した。

また、レビュー文書全体を読まずに済ませるために、短い要約を生成する研究が行われている。Glaserら[51]は、本文と同一の極性とその根拠 (ドメインに類出する名詞句) を含む文を抽出し、Amazonのレビューの要約とする。評価は、Mechanical Turkを用いて他の抽出手法 (極性の一番強い文) と比べてどちらが根拠を含んでいるか比較している。Ganesanら[90]は、アスペクト表現と意見表現を組み合わせた単語列を、要約として抽出した。要約は、頻出語をシードとして、元の文書に対する相互情報量により意見の代表性のスコアを計算し、n-グラム言語モデルに基づき語順のスコアを与えつつ、前後の単語を拡張することにより生成する。評価コーパスは、CNET^{*49}から収集した商品レビュー^{*50}を対象として、2人のアノテータが、レビューサイトが提供している賛否意見を参考にして、理想的な要約を作成した。また、ROUGE[93]やQuality Question^{*51}などオーソドックスな要約の評価を行った。

最後に、「昼間なら良い写真が撮れる」というレビュー文における「昼間なら」のような関連する文脈の損失

を避けるために、意見の周囲の修飾要素も含めてグラフ構造として抽出する研究がある。Wuら[91]は、意見表現とアスペクト表現の関係を (対象) により、意見表現と修飾語の関係を (制約, 展開) により、意見表現動詞の関係を (共起, 遷移) により、それぞれエッジとして連結することで、文を入力としてグラフ表現を出力する意見マイニング手法を提案した。評価コーパスは、中国語の商品レビューを対象として、2人のアノテータが、意見表現とそれに対する修飾や対象などの関係表現をアノテートした。

3.3 ブログを主な対象とした研究

ブログを対象とした意見分析の研究は、下火となっており、今回調査した中では2件[26],[64]だけであったが、世論の分析への応用を目的として、単独のブログポストではなく、ブログフィード (ブログによる同一サイト上のブログの集合) を対象とした意見分析や、感情を判別する研究が行われている。

Xuら[64]は、2.3節で紹介したTREC Blog08テストコレクションを利用して、与えられたクエリに対して、クエリについての意見表現を潜在変数として、ブログフィードを生成する確率に基づき、クエリの話題に適合した意見を表明しているブログフィードをランク付けした。意見性の判別の評価については、プーリング[94]に基づく正解データがTRECから提供されている。一方、Socherら[26]は、個人のネット上の体験を、深層学習の機械学習器であるautoencoderの出力を再帰的にマージする半教師あり学習に基づき、“お悔やみ”、“賞賛”、“楽しい逸話”、“共感”、“衝撃”に分類した。評価コーパス^{*52}は、個人の体験を告白するexperienceプロジェクト^{*53}を対象とした。各カテゴリは、サイトの閲覧者の投票の多数決により決定されている。

3.4 マイクロブログを対象とした研究

マイクロブログ、中でも特にツイッター^{*54}を対象とした意見分析の研究は急速にさかんになっている[95]。つぶやきには他のSNS同様に、意見を含むものが多くあるが、一方でつぶやきの文字数制限のために、単体のつぶやきを対象として極性などを分類することは難しい。そのため、つぶやきの文脈やTwitter特有の手がかりを利用して極性分類の精度を上げる研究がさかんに行われている。ここでは、5つの研究[17],[96],[97],[98],[99]について紹介する。

Jiangら[96]は、レビューと異なり、Twitterの話題は多岐にわたることから、すべての意見ではなく、クエリを

^{*47} <http://www.imdb.com>

^{*48} <http://www.cs.jhu.edu/~mdredze/datasets/sentiment/>

^{*49} <http://www.cnet.com>

^{*50} <http://timan.cs.uiuc.edu/downloads.html>

^{*51} <http://duc.nist.gov/duc2005/quality-questions.txt>

^{*52} <http://www.socher.org/index.php/Main/Semi-SupervisedRecursiveAutoencodersForPredictingSentimentDistributions>

^{*53} <http://www.experienceproject.com/confessions.php>

^{*54} <http://twitter.com>

意見の対象として含むもの、具体的には、対象と構文関係を持つ意見表現について、分析する手法を提案した。さらに、Twitterが短い点を考慮し、関連するツイート（同一人物のツイート、リプライ、リツイート）を考慮に入れ、極性を分類した。評価コーパスは、5つのクエリを準備し、各クエリについて400のつぶやきを収集し、クエリを意見の対象とした場合の極性を、著者らが分担して付与し、正解データとして評価を行った。判定の一貫性を調査するために、2人のアノテータにより400のうち100のつぶやきを判定した結果、肯定と否定との分類は一致するのに対して、極性なしと極性ありとの分類は一致しないものの、全体としては86%の判定が一致することを明らかにした。一方、Wangら[98]は、ある話題についての意見を収集するために、Twitter特有のメタデータであり、つぶやきの内容を表すハッシュタグに着目した。さらに、あるハッシュタグとつぶやき上で共起するハッシュタグは、極性を共有するという観察に基づき、共起するハッシュタグおよびそのハッシュタグに含まれる極性語を素性として、ハッシュタグが表す話題の極性を判別した。評価は、強い肯定・否定の極性語を文字列として含むハッシュタグを正解データとしている。Brodyら[99]は、マイクロブログ特有のインフォーマルな表記に着目し、自らの意見を強調するために、意見表現（“nice”, “love” など）の中の同じ文字が繰り返されており、これらの表現がつぶやきの極性判定に有用なことを明らかにしている。評価コーパスは、50万件的ツイートを米国周辺などの地理情報に基づき収集し、つぶやきを対象としたアノテーションは行っていない。

Purverら[17]は、つぶやきを6つの感情に分類するために、顔文字とハッシュタグを手がかりとした手法を提案した。また、3人のアノテータ間で、つぶやきを感情に分類した結果、一致率は、“恐れ”、“驚き”、“嫌悪”について特に低いことが分かった。評価コーパスは、3人とも同じ感情に分類したものだけを正解データとした。Silvaら[97]は、マイクロブログのストリームデータを対象とした意見分析手法を提案し、ブラジルの大統領選挙における候補者への支持、TIME紙の今年の人々の発表に対する感情、FIFAワールドカップにおけるブラジル代表選手の選出に対する感情について、3-5人のアノテータが手動で付与したコーパスに基づき評価している。

3.5 その他

Kucuktuncら[100]は、コミュニティQAデータを対象とした意見分析について取り組んでいる。この論文では、カテゴリ間の意見の性質の差、性差、年齢差、時間帯の差、経験による差、ベストアンサーにおける差など、コミュニティQAの分析に研究の重点が行われている。意見分析自体は、汎用的なシステムを利用している。また、SNSユーザ間の意見の交流に基づく影響の発見[101]なども行われ

ており、ユーザ間の交流への応用が今後はさかんになってくると考えられるが、こうした点に重点をおいた意見分析コーパスの開発は、今後の課題と考えられる。

そのほかには、別の文書ジャンルを対象とした応用[102], [103], [104], [105]や、辞書の構築[106], [107], [108], [109]について、研究が行われている。

以上から、意見分析についての最近の研究では、ニュース文書については、既存の意見分析コーパスを積極的に活用しているのに対して、レビュー文書を対象とした場合には、実際にレビューを活用している顧客ユーザの情報要求を考慮した新たなコーパスを構築することが多い。ブログについては、顧客ユーザが書き込む目的がはっきりしているレビューとは異なり、多様な話題についての書き込みが自由に許されるという点で、文書ジャンルの性質としてはニュース文書に近いと考えられるが、CGMを対象とした応用に共通する性質として、多種多様かつ大量のブログを収集したコーパスに価値があることが多く、広く活用されているコーパスが少ない。一方で、マイクロブログやコミュニティQAなどのSNSデータについては、コーパスの必要性や構築方針が検討されている段階であるといえる。

4. 意見分析コーパス構築の課題解決に向けた取り組み

本章では、これまでの章で見てきた傾向をふまえて、意見分析コーパスを構築するにあたっての課題に取り組んでいる研究を紹介する。まず、アノテータによりラベルつきコーパスを構築する際に考慮する点について述べられている研究を紹介した後で、手がかりを用いて擬似ラベルつきコーパスを自動的に構築する研究について紹介する。

4.1 アノテータによる意見分析コーパスの構築

アノテータにより意見分析コーパスを構築する初期の研究としては、Wiebeらの研究[9]がある。この研究では、アノテータ間のアノテーションのずれの解消という観点から、以下の手順により、アノテーションが行われている。

- (1) 教師あり学習への応用を目的として、4人のアノテータが、*Wall Street Journal*の記事における単文か節を単位として主観性・客観性のラベルをアノテーションを行う。
- (2) 4人の付与した情報を手がかりとして潜在変数を推定し、各アノテータのバイアスが修正されたラベルを提示し、アノテータ間で議論をしておし、必要があればアノテーションとマニュアルを修正した後、別のコーパスを対象として、アノテーションを行う。

なお、4人のアノテータのうち1人が途中の議論に参加しなかったが、この場合、この1人についての2回目のアノテーションの一致率(κ 係数)は向上しなかったことが明らかとなっている。その他の3人については、 κ 係数は

上記の手順により、0.7-0.8程度の値が得られている。この研究からは、以下のことが考察できる。

- ニュースを対象とした場合、アノテータが意見性を正確に判定するためには、アノテータの訓練が必要である。特に、複数のアノテータ間のずれに対して、議論をして判定基準を安定させることが重要である。
- アノテータ間の一致率を評価する尺度として、 κ 係数を用いることができる。アノテータの訓練を終えた後であれば、0.7-0.8程度の κ 値が得られるとよい。
- 訓練が初期段階のアノテーションから自動推定した擬似的なラベルにより、判定者自身の誤りを気づかせることができる。これは、多くのアノテータが誤りを含みつつも付与したラベルから、正解ラベルを推定できる可能性を示唆している。

これは人手で意見分析コーパスを構築するための手順として、妥当なやり方と考えられ、現在、ニュースを対象とした意見分析コーパスを構築するうえでは、 κ 係数により、アノテータ間の判定一致率を調査することは慣例となっている。しかし、上記のような手続きは、コスト（アノテータ自身の労力、アノテータに対して支払う金額、コーパスを構築させるまでの時間）を避けることは難しい。この問題に対して、アノテータの訓練を考慮しない方法として、クラウドソーシングに基づくアノテーション [14] が提案された。Snowら [14] は、(Amazon) Mechanical Turk^{*55}を用いて、訓練を受けた専門家とは異なるコストの少ないアノテータ（クラウドソース上のネットユーザ）から、自然言語処理において有用なコーパスが得られることを主張している。この研究において注意すべき点は、クラウドソース上のユーザに行わせるタスクを、自然かつ非専門家でも学習可能なタスクとしているところにある。具体的には、複数の項目からの選択と、ある範囲内の数値の入力のみ制限しており、テキスト中の文字列の抽出などは対象としていない。また、アノテーション方針は、Webを通じてアクセス可能な状態にしてある。文献 [9] とのアノテーションにおける大きな違いは、以下の2点にある。

- タスク設計者とアノテータの間で、アノテーション方針に関する議論を行わない。
- 多く（各判定項目あたり10人以上）のアノテータを、安価で雇用できる。

彼らが対象としたタスクは5つあるが、意見分析に関係しているタスクとしては、2.1節で紹介した、*SemEval-2007 Task 14: Affective Text* [38]^{*56}におけるアノテーションを取り上げている。これは、ニュースの見出しを対象として、“怒り”、“嫌悪”、“恐れ”、“快楽”、“悲しみ”、“驚き”の6つの感情について [0-100] の間の値を割り当て、一方で肯定から否定の間を、[-100-100] の間の値を割り当てるとい

ものである。このタスクについて、アノテータ間の一致率を、ピアソンの相関係数を用いて評価した。すると、訓練を受けたアノテータ間の一致率は、クラウドソース上のアノテータ間の一致率を上回っているものの、クラウドソース上のアノテータの付与した値の平均値を1人のアノテータと見なした場合、“恐れ”を除いて2-9人（平均4人）の平均値で、訓練を受けたアノテータ1人よりも、他のアノテータとの相関が高くなることを明らかにした。また、少数の正解データとの判定のずれに基づき、ずれが少なくなるほど重みが大きくなるように、クラウドソース上のアノテータの判定を重み付けした場合、相関がわずかではあるが一貫して向上することを明らかにした。

そのほか、手動で構築したコーパスを拡張する方法として、ラベルのサイズ、たとえば肯定文書数と否定文書数が不均衡な意見分析コーパスについて、能動学習により、少数のラベルの文書について、極性分類に有用なサンプルを選択し、アノテーションを支援する研究がある [59]。評価コーパスは、先行研究 [110] で用いた、Blitzerら [53] と同様の4つのドメインを対象としたレビュー文書を用いている。この手法により、人手によるラベル付けの労力を最小に保ちつつ、分類に有効な訓練データを作成することが期待できる。

4.2 手がかりを用いた擬似ラベルつきコーパスの自動構築

アノテータを用いずにラベルつきコーパスを作成した初期の研究として、Pangら [42] があげられる。2.2節でもコーパスを紹介したとおり、映画サイトのレビューは、サイトの投稿者により5つ星を単位とした評価がつけられているため、教師あり学習に基づき極性を分類しようとした場合に、訓練データを人手で構築する必要がない。また、手がかりを用いてコーパスを構築する初期の研究としては、Kajiら [15] があげられる。彼らは、レビューサイトにおける極性データについて、文書を単位としているため、文書内に良い点も悪い点も含まれていた場合、全体で「悪い」と評価されている場合に「良い」点がノイズとなるという問題を指摘した。具体的には、“この映画のあらすじは良く、俳優もキャストも一級品だが、見る価値はない。”というレビューについて、文書全体で否定的なレビューとしてしまうと、“あらすじが良い”、“俳優もキャストも一級品”という評価はノイズとなる。このため、Web文書から表層的な手がかりを用いて、肯定（否定）意見を、文を単位として抽出する手法を提案した。Web文書から“賛成意見”、“反対意見”などの言葉が見出しにあり、その後の項目に、●などで始まる箇条書きが続く場合、その箇条書きの中には“賛成（反対）意見”が含まれると仮定できる。また、HTMLのテーブル構造の、行（列）見出しとして“賛成（反対）意見”が存在する場合に、右の列（下の行）には、具体的な賛成（反対）意見が記述されているとも仮定

^{*55} <http://mturk.com>

^{*56} <http://www.cs.unt.edu/~rada/affectivetext>

できる。さらに，“利点（欠点）は…こと.” というような文があれば，“…”の部分には何らかの肯定（否定）意見が記述されていると考えられる。これらの HTML に基づく箇条書きや表などのレイアウト構造，あるいは，手がかり語を利用した表層パターンに基づき，極性を付与した意見を抽出し，コーパスを構築した。

一方，マイクロブログを対象とした場合，極性を分類するための意見分析コーパスを自動構築する手がかりは，2.3 節でも述べたとおり，顔文字が利用される [70]。さらに，極性にとどまらず，顔文字などを手がかりとして，感情のような詳細なカテゴリを分類したコーパスを自動的に構築する研究 [16], [17] も行われている。Ptaszynski らの研究 [16] は，約 1,300 万件（3 億文以上）の日本語のブログを対象として，10 の感情を自動的に付与することを試みており，1,000 件のサンプルに対する分類精度は約 90% と，非常に高いことが報告されている。3.4 節でも紹介した Purver ら [17] の研究は，英語のマイクロブログを対象としており，顔文字とハッシュタグを手がかりとして，6 つの感情の分類に取り組んだ結果，3 つの感情（幸福，悲しみ，怒り）についてのみ区別が明確（分類精度は 80% 弱）となった。この結果を検証するために，*Mechanical Turk* を使用して人手でラベル付けを行い，アノテータ間の一致率を調べた結果，残りの 3 つの感情（驚き，恐れ，嫌悪）については一致率が低いことが分かった。ここから分かることとしては，以下のことがあげられる。

- 手動で一貫して付与が行えないラベルは，コーパスの自動構築に利用できない。したがって，詳細カテゴリを使用したコーパスを自動構築するためには，まず手動で一貫してラベルを付与できるカテゴリを対象とする必要がある。
- ブログとマイクロブログの傾向の差異については，文書ジャンルごとに利用できる手がかりの違いの影響があると考えられる。

また，別のアプローチとして，以前は，意見分析のためのリソース，すなわちコーパスや辞書は，英語の整備状況が進んでおり，日本語は例外的に整備が進んでいたものの，他の言語のリソースは貧弱であった。このような状況においては，英語のリソースを他言語に翻訳し活用するアプローチが提案されていた [111]。その後，3.1 節でも紹介したとおり，機械翻訳において語彙が制約される問題を考慮し，MPQA 意見コーパスと NTCIR 多言語意見分析コーパスを訓練データとして使用し，対訳コーパスから極性データを自動的に構築し，両言語のリソースを使用することにより極性分類の精度を向上させる研究 [28] も行われている。

5. おわりに

本稿では，言語処理研究で，利用されることの多い意見分析コーパスを，文書ジャンル別に紹介し，各ジャンル

における意見分析コーパスの基本的な要件を明らかにした。また，最近の意見分析に関する研究を対象として，文書ジャンルと研究目的に応じてアノテートする必要のある属性や，コーパスを活用することにより可能となる応用研究を明らかにした。さらに，コーパス構築にあたっての取り組みについて，アノテータを雇用する場合と，自動的に擬似ラベルデータを構築する場合とに分けて整理して紹介し，アノテーションのクオリティを維持しつつ，コストを下げるための工夫や，アノテーションのカテゴリを極性だけではなく感情に拡張する研究について紹介した。これらの研究は，文書ジャンルごとの性質への依存性が見られ，今後，公開される意見分析コーパスが増えることにより，さらなる取り組みが期待される。また，SNS，マイクロブログ，コミュニティ QA など，利用者が意見を発信するメディアが発展するにともない，意見分析研究およびそのためのコーパス開発の重要性は引き続き増していくと考えられる。

謝辞 本研究の一部は，科学研究費補助金基盤研究 C（課題番号 24500291），基盤研究 B（課題番号 25280110）ならびに萌芽研究（課題番号 25540159）の助成を受けて遂行された。

貴重なコメントをいただいたメタレビューアならびに査読者の方々に深謝いたします。

参考文献

- [1] Pang, B. and Lee, L.: *Opinion Mining and Sentiment Analysis*, Now Publishers Inc. (2008).
- [2] Liu, B.: Sentiment Analysis and Subjectivity, *Handbook of Natural Language Processing*, Indurkha, N. and Damerau, F.J. (Eds.), pp.627-664, CRC Press (2010).
- [3] 大塚裕子，乾 孝司，奥村 学：意見分析エンジン，コロナ社 (2007)。
- [4] 乾 孝司，奥村 学：テキストを対象とした評価情報の分析に関する研究動向，自然言語処理，Vol.13, No.3, pp.201-241 (2006)。
- [5] Wiebe, J., Wilson, T. and Cardie, C.: Annotating Expressions of Opinions and Emotions in Language, *Language Resources and Evaluation*, Vol.39, No.2-3, pp.165-210 (2005)。
- [6] 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，情報処理学会第 171 回自然言語処理研究会，pp.111-118 (2006)。
- [7] Bazerman, C. and Prior, P.: *What Writing Does and How It Does It*, Lawrence Erlbaum Associates (2004)。
- [8] Stoyanov, V. and Cardie, C.: Automatically Creating General-Purpose Opinion Summaries from Text, *Proc. Recent Advances in Natural Language Processing (RANLP 2011)*, Hissar, Bulgaria, pp.202-209 (2011)。
- [9] Wiebe, J., Bruce, R. and O'Hara, T.P.: Development and Use of a Gold Standard Data Set for Subjectivity Classifications, *Proc. 37th Annual Meeting of the Assoc. for Computational Linguistics (ACL-99)*, pp.246-253, University of Maryland (1999)。
- [10] 玉木 明：ニュース報道の言語論，洋泉社 (1996)。

- [11] Feldman, R.: Techniques and Applications for Sentiment Analysis, *Comm. ACM*, Vol.56, No.4, pp.82-89 (2013).
- [12] Carenini, G., Ng, R. and Pauls, A.: Multi-Document Summarization of Evaluative Text, *Proc. 11th European Chapter of the Association for Computational Linguistics (EACL)*, Trento, Italy, pp.305-312 (2006).
- [13] Yu, J., Zha, Z.-J., Wang, M., Wang, K. and Chua, T.-S.: Domain-Assisted Product Aspect Hierarchy Generation: Towards Hierarchical Organization of Unstructured Consumer Reviews, *Proc. 2011 Conf. on Empirical Methods in Natural Language Processing (EMNLP 2011)*, Edinburgh, Scotland, UK, pp.140-150 (2011).
- [14] Snow, R., O'Connor, B., Jurafsky, D. and Ng, A.Y.: Cheap and Fast? - But is it Good? Evaluating Non-Expert Annotations for Natural Language Tasks, *Proc. 2008 Conf. on Empirical Methods in Natural Language Processing (EMNLP 2008)*, Waikiki, Honolulu, Hawaii, pp.254-263 (2008).
- [15] Kajji, N. and Kitsuregawa, M.: Automatic Construction of Polarity-tagged Corpus from HTML Documents, *Proc. 21st International Conf. on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics (COLING-ACL 2006)*, Sydney, Australia, pp.452-459 (2006).
- [16] Ptaszynski, M., Rzepka, R., Araki, K. and Momouchi, Y.: Automatically Annotating A Five-Billion-Word Corpus of Japanese Blogs for Affect and Sentiment Analysis, *Proc. 3rd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA)*, Jeju, Republic of Korea, pp.89-98 (2012).
- [17] Purver, M. and Battersby, S.: Experimenting with Distant Supervision for Emotion Classification, *Proc. 13th European Chapter of the Association for Computational Linguistics (EACL 2012)*, Avignon, France, pp.482-491 (2012).
- [18] Wiebe, J., Breck, E., Buckley, C., Cardie, C., Davis, P., Fraser, B., Litman, D., Pierce, D., Riloff, E. and Wilson, T.: NRRC Summer Workshop on Multiple-Perspective Question Answering Final Report (2002).
- [19] Cohen, J.: A Coefficient of Agreement for Nominal Scales, *Educational and Psychological Measurement*, Vol.20, No.1, pp.37-46 (1960).
- [20] Cardie, C., Wiebe, J., Wilson, T. and Litman, D.: Combining Low-Level and Summary Representations of Opinions for Multi-Perspective Question Answering, *Proc. AAAI Spring Sympo. on New Directions in Question Answering*, Stanford, CA, pp.20-27 (2003).
- [21] Choi, Y., Cardie, C., Riloff, E. and Patwardhan, S.: Identifying Sources of Opinions with Conditional Random Fields and Extraction Patterns, *Proc. 2005 Human Language Technology Conf. and Conf. on Empirical Methods in Natural Language Processing (HLT/EMNLP 2005)*, Vancouver, B.C. (2005).
- [22] Wilson, T., Wiebe, J. and Hoffmann, P.: Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis, *Proc. 2005 Human Language Technology Conf. and Conf. on Empirical Methods in Natural Language Processing (HLT/EMNLP 2005)*, Vancouver, B.C. (2005).
- [23] Stoyanov, V., Cardie, C. and Wiebe, J.: Multi-Perspective Question Answering Using the OpQA Corpus, *Proc. 2005 Human Language Technology Conf. and Conf. on Empirical Methods in Natural Language Processing (HLT/EMNLP 2005)*, Vancouver, B.C. (2005).
- [24] Yessenalina, A. and Cardie, C.: Compositional Matrix-Space Models for Sentiment Analysis, *Proc. 2011 Conf. on Empirical Methods in Natural Language Processing (EMNLP 2011)*, Edinburgh, Scotland, UK, pp.172-182 (2011).
- [25] Wiegand, M. and Klakow, D.: Generalization Methods for In-Domain and Cross-Domain Opinion Holder Extraction, *Proc. 13th Conf. of the European Chapter of the Association for Computational Linguistics (EACL 2012)*, Avignon, France, pp.325-335 (2012).
- [26] Socher, R., and Eric H. Huang, J.P., Ng, A.Y. and Manning, C.D.: Semi-Supervised Recursive Autoencoders for Predicting Sentiment Distributions, *Proc. 2011 Conf. on Empirical Methods in Natural Language Processing (EMNLP 2011)*, Edinburgh, Scotland, UK, pp.151-161 (2011).
- [27] Yang, B. and Cardie, C.: Extracting Opinion Expressions with semi-Markov Conditional Random Fields, *Proc. 2012 Conf. on Empirical Methods in Natural Language Processing (EMNLP 2012)*, Jeju Island, Republic of Korea, pp.1335-1345 (2012).
- [28] Lu, B., Tan, C., Cardie, C. and Tsou, B.K.: Joint Bilingual Sentiment Classification with Unlabeled Parallel Corpora, *Proc. 49th Annual Meeting of the Association for Computational Linguistics (ACL 2011)*, Portland, Oregon, pp.320-330 (2011).
- [29] Lu, Y., Castellanos, M., Dayal, U. and Zhai, C.: Automatic Construction of a Context-Aware Sentiment Lexicon: An Optimization Approach, *Proc. 20th International Conf. on World Wide Web (WWW 2011)*, Hyderabad, India, pp.347-356 (2011).
- [30] Wiebe, J. and Mihalcea, R.: Word Sense and Subjectivity, *Proc. 21st International Conf. on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics (COLING-ACL 2006)*, Sydney, Australia, pp.1065-1072 (2006).
- [31] Somasundaran, S. and Wiebe, J.: Recognizing Stances in Online Debates, *Proc. 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conf. on Natural Language Processing of the Asian Federation of Natural Language Processing*, Singapore (2009).
- [32] Somasundaran, S. and Wiebe, J.: Recognizing Stances in Ideological On-line Debates, *Proc. NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, Los Angeles, CA, pp.116-124 (2010).
- [33] Seki, Y., Ku, L.W., Sun, L., Chen, H.H. and Kando, N.: Overview of Multilingual Opinion Analysis Task at NTCIR-8 - A Step Toward Cross Lingual Opinion Analysis, *Proc. 8th NTCIR Workshop Meeting*, NII, Japan, pp.209-220 (2010).
- [34] Seki, Y., Evans, D.K., Ku, L.W., Sun, L., Chen, H.H. and Kando, N.: Overview of Multilingual Opinion Analysis Task at NTCIR-7, *Proc. 7th NTCIR Workshop Meeting*, NII, Japan, pp.185-203 (2008).
- [35] Seki, Y., Evans, D.K., Ku, L.W., Chen, H.H., Kando, N. and Lin, C.Y.: Overview of Opinion Analysis Pilot Task at NTCIR-6, *Proc. 6th NTCIR Workshop Meeting*, NII, Japan, pp.265-278 (2007).
- [36] Zhou, L., Li, B., Gao, W., Wei, Z. and Wong, K.-F.: Unsupervised Discovery of Discourse Relations for Eliminating Intra-sentence Polarity Ambiguities, *Proc. 2011 Conf. on Empirical Methods in Natural Language Processing (EMNLP 2011)*, Edinburgh, Scotland, UK, pp.162-171 (2011).

- [37] Meng, X., Wei, F., Liu, X., Zhou, M., Xu, G. and Wang, H.: Cross-Lingual Mixture Model for Sentiment Classification, *Proc. 50th Annual Meeting of the Association for Computational Linguistics (ACL 2012)*, Jeju, Republic of Korea, pp.572–581 (2012).
- [38] Strapparava, C. and Mihalcea, R.: SemEval-2007 Task 14: Affective Text, *Proc. 4th International Workshop on Semantic Evaluations (SemEval-2007)*, Prague, Czech Republic, pp.70–74 (2007).
- [39] Balahur, A. and Steinberger, R.: Rethinking Sentiment Analysis in the News: From Theory to Practice and Back, *Proc. 1st Workshop on Opinion Mining and Sentiment Analysis (WOMSA)*, Sevilla, Spain, pp.1–12 (2009).
- [40] Balahur, A., Steinberger, R., Kabadjov, M., Zavarella, V., van der Goot, E., Halkia, M., Pouliquen, B. and Belyaeva, J.: Sentiment Analysis in the News, *Proc. 7th International Conf. on Language Resources and Evaluation (LREC 2010)*, Valletta, Malta, pp.2216–2220 (2010).
- [41] Steinberger, R., Pouliquen, B. and van der Goo, E.: An introduction to the Europe Media Monitor, *Proc. ACM SIGIR 2009 Workshop: Information Access in a Multilingual World*, Boston, MA, USA (2009).
- [42] Pang, B., Lee, L. and Vaithyanathan, S.: Thumbs up? Sentiment Classification using Machine Learning Techniques, *Proc. 2002 Conf. on Empirical Methods in Natural Language Processing (EMNLP 2002)*, pp.79–86 (2002).
- [43] Pang, B. and Lee, L.: A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts, *Proc. 42nd Ann. Meeting of the Assoc. for Computational Linguistics (ACL 2004)*, Barcelona, Spain, pp.271–278 (2004).
- [44] Pang, B. and Lee, L.: Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales, *Proc. 43rd Annual Meeting of the Association for Computational Linguistics (ACL 2005)*, Ann Arbor, Michigan, pp.115–124 (2005).
- [45] Maas, A.L., E. Daly, R., Pham, P.T., Huang, D., Ng, A.Y. and Potts, C.: Learning Word Vectors for Sentiment Analysis, *Proc. 49th Annual Meeting of the Association for Computational Linguistics (ACL 2011)*, Portland, Oregon, pp.142–150 (2011).
- [46] Hu, M. and Liu, B.: Mining and Summarizing Customer Reviews, *Proc. 10th ACM SIGKDD International Conf. on Knowledge Discovery and Data Mining (KDD 2004)*, Seattle, Washington, USA (2004).
- [47] Ding, X., Liu, B. and Yu, P.S.: A Holistic Lexicon-Based Approach to Opinion Mining, *Proc. 1st ACM International Conf. on Web Search and Data Mining (WSDM 2008)*, Stanford University, California, USA (2008).
- [48] Jindal, N. and Liu, B.: Opinion Spam and Analysis, *Proc. 1st ACM International Conf. on Web Search and Data Mining (WSDM 2008)*, Stanford University, California, USA (2008).
- [49] Jindal, N. and Liu, B.: Identifying Comparative Sentences in Text Documents, *Proc. 29th Annual International ACM SIGIR Conf. on Research and Development on Information Retrieval (SIGIR 2006)*, Seattle, USA (2006).
- [50] Ganapathibhotla, M. and Liu, B.: Mining Opinions in Comparative Sentences, *Proc. 22nd International Conf. on Computational Linguistics (COLING 2008)*, Manchester (2008).
- [51] Glaser, A. and Schutze, H.: Automatic generation of short informative sentiment summaries, *Proc. 13th Conf. of the European Chapter of the Association for Computational Linguistics (EACL 2012)*, Avignon, France, pp.276–285 (2012).
- [52] Mukherjee, A., Liu, B. and Glance, N.: Spotting Fake Reviewer Groups in Consumer Reviews, *International World Wide Web Conf. (WWW 2012)*, Lyon, France (2012).
- [53] Blitzer, J., Dredze, M. and Pereira, F.: Biographies, Bollywood, Boom-boxes and Blenders: Domain Adaptation for Sentiment Classification, *Proc. 45th Annual Meeting of the Association of Computational Linguistics (ACL 2007)*, Prague, Czech Republic, pp.440–447 (2007).
- [54] He, Y., Lin, C. and Alani, H.: Automatically Extracting Polarity-Bearing Topics for Cross-Domain Sentiment Classification, *Proc. 49th Annual Meeting of the Association for Computational Linguistics (ACL 2011)*, Portland, Oregon, pp.123–131 (2011).
- [55] Bollegala, D., Weir, D. and Carroll, J.: Using Multiple Sources to Construct a Sentiment Sensitive Thesaurus for Cross-Domain Sentiment Classification, *Proc. 49th Annual Meeting of the Association for Computational Linguistics (ACL 2011)*, Portland, Oregon, pp.132–141 (2011).
- [56] Li, F., Pan, S.J., Jin, O., Yang, Q. and Zhu, X.: Cross-Domain Co-Extraction of Sentiment and Topic Lexicons, *Proc. 50th Annual Meeting of the Association for Computational Linguistics (ACL 2012)*, Jeju, Republic of Korea, pp.410–419 (2012).
- [57] Ponomareva, N. and Thelwall, M.: Do Neighbours Help? An Exploration of Graph-based Algorithms for Cross-domain Sentiment Classification, *Proc. 2012 Conf. on Empirical Methods in Natural Language Processing (EMNLP 2012)*, Jeju Island, Republic of Korea, pp.655–665 (2012).
- [58] Beshpalov, D., Bai, B., Qi, Y. and Shokoufandeh, A.: Sentiment Classification Based on Supervised Latent n-gram Analysis, *Proc. 20th ACM Conf. on Information and Knowledge Management (CIKM 2011)*, Glasgow, Scotland, UK, pp.375–382 (2011).
- [59] Li, S., Ju, S., Zhou, G. and Li, X.: Active Learning for Imbalanced Sentiment Classification, *Proc. 2012 Conf. on Empirical Methods in Natural Language Processing (EMNLP 2012)*, Jeju Island, Republic of Korea, pp.139–148 (2012).
- [60] Hong, Y., Lu, J., Yao, J., Zhu, Q. and Zhou, G.: What Reviews are Satisfactory: Novel Features for Automatic Helpfulness Voting, *Proc. 35th international ACM SIGIR Conf. on Research and development in Information Retrieval (SIGIR 2012)*, Portland, Oregon, pp.495–504 (2012).
- [61] Taboada, M., Anthony, C. and Voll, K.: Methods for Creating Semantic Orientation Dictionaries, *Proc. 5th International Conference on Language Resources and Evaluation (LREC 2006)*, Genoa, Italy, pp.427–432 (2006).
- [62] Wang, H., Lu, Y. and Zhai, C.: Latent Aspect Rating Analysis on Review Text Data: A Rating Regression Approach, *Proc. 16th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD 2010)*, New York, NY, USA, pp.783–792 (2010).
- [63] Macdonald, C., Santos, R.L., Ounis, I. and Soboroff, I.:

- Blog Track Research at TREC, *ACM SIGIR Forum*, Vol.44, No.1, pp.58-75 (2010).
- [64] Xu, X., Tan, S., Liu, Y., Cheng, X., Lin, Z. and Guo, J.: Find Me Opinion Sources in Blogosphere: A Unified Framework for Opinionated Blog Feed Retrieval, *Proc. 5th International Conf. on Web Search and Web Data Mining (WSDM 2012)*, Seattle, WA, USA, pp.583-592 (2012).
- [65] Dang, H.T.: Overview of the TAC 2008 Opinion Question Answering and Summarization Tasks, *Proc. Text Analysis Conf. 2008 (TAC 2008)*, pp.24-35 (2008).
- [66] Lin, J. and Demner-Fushman, D.: Will Pyramids Built of Nuggets Topple Over?, *Proc. 2006 Human Language Technology Conf. and the North American Chapter of the Association for Computational Linguistics Annual Meeting (HLT/NAACL 2006)*, New York City, USA, pp.383-390 (2006).
- [67] Boldrini, E., Balahur, A., Martinez-Barco, P. and Montoyo, A.: EmotiBlog: A Finer-Grained and More Precise Learning of Subjectivity Expression Models, *Proc. 4th Linguistic Annotation Workshop (LAW 2010), ACL 2010*, Uppsala, Sweden, pp.1-10 (2010).
- [68] Boldrini, E., Fernandez, J., Gomez, J.M. and Martinez-Barco, P.: Evaluation the Robustness of EmotiBlog for Sentiment Analysis and Opinion Mining, *Proc. Recent Advances in Natural Language Processing (RANLP 2011)*, Hissar, Bulgaria, pp.521-526 (2011).
- [69] Zhao, J., Xu, H., Huang, X., Tan, S., Liu, K. and Zhang, Q.: Overview of Chinese Opinion Analysis Evaluation 2008 (in Chinese), *Proc. 1st Chinese Opinion Analysis Evaluation* (2008).
- [70] Go, A., Bhayani, R. and Huang, L.: Twitter Sentiment Classification using Distant Supervision, Technical report, CS224N Project Report, Stanford (2009).
- [71] Turney, P.: Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews, *Proc. 40th Annual Meeting on Association for Computational Linguistics*, Philadelphia, USA, pp.417-424 (2002).
- [72] Shanahan, J.G., Qu, Y. and Wiebe, J.M.: *Computing Attitude and Affect in Text: Theory and Applications*, Springer-Verlag (2006).
- [73] Breck, E., Choi, Y. and Cardie, C.: Identifying expressions of opinion in context, *Proc. 20th Int'l Joint Conference on Artificial Intelligence (IJCAI 2007)*, Hyderabad, India, pp.2683-2688 (2007).
- [74] Park, S., Lee, K.S. and Song, J.: Contrasting Opposing Views of News Articles on Contentious Issues, *Proc. 49th Annual Meeting of the Association for Computational Linguistics (ACL 2011)*, Portland, Oregon, pp.340-349 (2011).
- [75] Fang, Y., Si, L., Somasundaram, N. and Yu, Z.: Mining Contrastive Opinions on Political Texts using Cross-Perspective Topic Model, *Proc. 5th International Conf. on Web Search and Web Data Mining (WSDM 2012)*, Seattle, WA, USA, pp.63-72 (2012).
- [76] Blei, D.M., Ng, A.Y. and Jordan, M.I.: Latent Dirichlet Allocation, *J. Mach. Learn. Res.*, Vol.3, pp.993-1022 (2003).
- [77] Balamurali, A.R., Joshi, A. and Bhattacharyya, P.: Harnessing WordNet Senses for Supervised Sentiment Classification, *Proc. 2011 Conf. on Empirical Methods in Natural Language Processing (EMNLP 2011)*, Edinburgh, Scotland, UK, pp.1081-1091 (2011).
- [78] Hai, Z., Chang, K. and Cong, G.: One Seed to Find Them All: Mining Opinion Features via Association, *Proc. 21th ACM Conf. on Information and Knowledge Management (CIKM 2012)*, Maui, HI, USA, pp.255-264 (2012).
- [79] Liu, K., Xu, L. and Zhao, J.: Opinion Target Extraction Using Word-Based Translation Model, *Proc. 2012 Conf. on Empirical Methods in Natural Language Processing (EMNLP 2012)*, Jeju Island, Republic of Korea, pp.1346-1356 (2012).
- [80] Moghaddam, S. and Ester, M.: On the Design of LDA Models for Aspect-based Opinion Mining, *Proc. 21th ACM Conf. on Information and Knowledge Management (CIKM 2012)*, Maui, HI, USA, pp.803-812 (2012).
- [81] Jo, Y. and Oh, A.: Aspect and Sentiment Unification Model for Online Review Analysis, *Proc. 4th International Conf. on Web Search and Web Data Mining (WSDM 2011)*, Hong Kong, China, pp.815-824 (2011).
- [82] Zhai, Z., Liu, B., Xu, H. and Jia, P.: Clustering Product Features for Opinion Mining, *Proc. 4th International Conf. on Web Search and Web Data Mining (WSDM 2011)*, Hong Kong, China, pp.347-354 (2011).
- [83] Sauper, C., Haghighi, A. and Barzilay, R.: Content Models with Attitude, *Proc. 49th Annual Meeting of the Association for Computational Linguistics (ACL 2011)*, Portland, Oregon, pp.350-358 (2011).
- [84] Yu, J., Zha, Z.-J. and Chua, T.-S.: Answering Opinion Questions on Products by Exploiting Hierarchical Organization of Consumer Reviews, *Proc. 2012 Conf. on Empirical Methods in Natural Language Processing (EMNLP 2012)*, Jeju Island, Republic of Korea, pp.391-401 (2012).
- [85] Lin, C. and He, Y.: Joint sentiment/topic model for Sentiment Analysis, *Proc. 18th ACM Conference on Information and Knowledge Management (CIKM 2009)*, Hong Kong, China, pp.375-384 (2009).
- [86] Li, F., Han, C., Huang, M., Zhu, X., Xia, Y.-J., Zhang, S. and Yu, H.: Structure-Aware Review Mining and Summarization, *Proc. 23rd International Conf. on Computational Linguistics (COLING 2010)*, Beijing, China, pp.653-661 (2010).
- [87] Ott, M., Choi, Y., Cardie, C. and Hancock, J.T.: Finding Deceptive Opinion Spam by Any Stretch of the Imagination, *Proc. 49th Annual Meeting of the Association for Computational Linguistics (ACL 2011)*, Portland, Oregon, pp.309-319 (2011).
- [88] Kawamae, N.: Predicting Future Reviews: Sentiment Analysis Models for Collaborative Filtering, *Proc. 4th International Conf. on Web Search and Web Data Mining (WSDM 2011)*, Hong Kong, China, pp.605-614 (2011).
- [89] Moshfeghi, Y., Piwowarski, B. and Jose, J.M.: Handling Data Sparsity in Collaborative Filtering using Emotion and Semantic Based Features, *Proc. 34th international ACM SIGIR Conf. on Research and development in Information Retrieval (SIGIR 2011)*, Beijing, China, pp.625-634 (2011).
- [90] Ganesan, K., Zhai, C.X. and Viegas, E.: Micropinion Generation: An Unsupervised Approach to Generating Ultra-Concise Summaries of Opinions, *Proc. 21st International Conf. on World Wide Web (WWW 2012)*, Lyon, France, pp.869-878 (2012).
- [91] Wu, Y., Zhang, Q., Huang, X. and Wu, L.: Structural Opinion Mining for Graph-based Sentiment Represent-

