

日本語意見分析コーパスの構築に向けた アスペクト情報の特徴分析

村上 浩司^{1,2} 中山 祐輝² Ikuko Hardaway¹

¹Rakuten Institute of Technology, Rakuten USA

² 楽天技術研究所 楽天グループ株式会社

{koji.murakami,yuki.b.,ikuko.hardaway}@rakuten.com

概要

我々はアスペクト情報を考慮した日本語意見分析コーパスを構築してきた。文中に存在するアスペクト表現を抽出するため、それらの認定に関してはその範囲の決定や対象表現の選別など、作業を進める上で考慮すべき点が多数ある。本論文ではアスペクト表現と評価表現について定義を行い、実際の小規模のアノテーション作業で明らかになった課題について整理すると共に、現状のコーパスを利用したアスペクト表現抽出の予備実験を行った。

1 はじめに

近年、製品に対するカスタマーレビューやツイートなどの意見を表現するテキストの分析の重要性が増している。文を対象とした肯定・否定の極性判定だけではなく、さらに具体的な情報を利用することで顧客の意図や要望を特定し利用することが求められている。例えばあるホテルの朝食について、「朝食の味噌汁は冷めていた」から「評価対象=朝食, 属性=品質, 評価表現(極性)=冷めていた(否定)」のような情報を抽出することで、文全体が否定であるという文の極性判定だけでは見えてこない業務の問題点などが明確化する。

そのためには評価極性だけではなく、評価対象、属性といったアスペクト情報を考慮した意見分析 (Aspect-Based Sentiment Analysis: ABSA) が重要であり [1, 2, 3, 4, 5], タスクを進める上でそれらの情報が付与されたコーパスが必須である。英語のデータセットはこれまでに多く開発され、そのうち SemEval で行われた ABSA タスクの評価コーパス [6, 7, 8] や MAMS[2] などが幅広く使われている。

日本語では NTCIR 多言語意見分析コーパス¹⁾や

TSUKUBA コーパス²⁾などウェブ上で入手できるものがいくつかあるが、これらはアスペクト情報を考慮しない意見分析に特化している [9, 10, 11, 12]。

新里ら [1] は、楽天市場の店舗レビューを“配送”、“梱包”などのアスペクトカテゴリとその極性カテゴリに分類する手法を提案したが、実験で用いられたコーパスは非公開である。栗原ら [13] は、ドメインを限定しない Twitter による評判分析を目的とした評価対象と評価表現を抽出するための日本語 ABSA データセット³⁾を構築した。論文化されていないものの、アスペクトカテゴリとその極性の評価に対応しているコーパスとして、上場企業の決算報告書をドメインとした chABSA-dataset⁴⁾がある。

意見分析タスクは文の評価対象などのアスペクト情報などを扱う方向性とは別に、極性の強さや感情の種類、意見の立場なども注目されており、近年こうした種類の情報を付与したコーパスも構築されている [14, 15]。

我々はアスペクト情報を考慮した日本語意見分析コーパスを構築してきた [17, 16]。このコーパスは楽天トラベル⁵⁾の施設に対するユーザーレビュー中の各文に対しアスペクトカテゴリ及びその評価極性の付与作業を行ったものである。本論文では現状のコーパスを拡張しアスペクト表現および評価表現を付与を目的として、そのために必要なコーパス設計と事例分析について述べる。具体的には抽出する情報の定義を行い、“朝食”が対象になる文を対象に先行して行ったアノテーション作業から、議論が必要な事例の分析と作業の方針を策定する。また、作業済みの文を対象にした2種類の ABSA 手法による予備抽出実験を行い、その結果から事例の特徴を分析

2) <https://www.nii.ac.jp/dsc/ldr/rakuten/>

3) http://www.cl.ecei.tohoku.ac.jp/resources/twitter_target_review

4) <https://github.com/chakki-works/chABSA-dataset/>

5) travel.rakuten.co.jp

1) <http://research.nii.ac.jp/ntcir/data/data-ja.html>

表1 アスペクト表現と評価極性に関する SemEval との比較

	SemEval2014[6]	SemEval2015[7]	SemEval2016 [8]	Nakayama[16]	本論文
アスペクトカテゴリ	✓	✓ (E#A)	✓ (E#A)	✓	(✓)
アスペクトカテゴリ極性	✓	✓	✓	✓	(✓)
アスペクト表現	✓	(OTE)	(OTE)		✓
アスペクト表現極性	✓				✓
評価表現					✓

することでコーパスの性質について考察する。

2 コーパスの設計と構築

これまでに構築した日本語意見分析コーパス [17, 16] を拡張し、アスペクト表現と評価表現を付与する。ここではアスペクト表現、評価表現の扱いと定義、作業方針について述べる。

2.1 アスペクト表現・評価表現の取り扱い

アスペクト表現は 2014 年の SemEval サブタスク (1)[6] において認識対象とされた。固有名詞、一般名詞、それらの複合名詞が認識すべきアスペクト表現となった。またアスペクト表現のクラス (レストランドメインでは、“食べ物”, “サービス”, “値段”などを予め定義) をアスペクトカテゴリとした。2015 年, 2016 年においてはアスペクト表現は直接の抽出対象とはなっておらず, アスペクトカテゴリの定義が変更された。新しいアスペクトカテゴリはエンティティ (E) とその属性 (A) の組み合わせ “E#A” で定義され, このアスペクトカテゴリとその極性認識がタスクとなった。E と A は予め定義されたエンティティタイプおよび属性ラベルからそれぞれ選択される。そしてさらに, エンティティの実体を指す表現 (OTE:Opinion Target Expressions) の抽出もタスクとなっている。また, SemEval においては一貫して評価表現の抽出については行っておらず, アスペクトカテゴリを同定する上で間接的に評価表現やその極性が利用されていると考えられる。

しかしながらこうした SemEval のタスク設定では, 書き手が「どんな種類の¹ことについてどういう態度で言及したか」は判定できるものの, 具体的に何についてどんな表現で満足・不満足を示しているかわからないことから, さらに我々は記載される実際のアスペクトを示す表現と評価表現の抽出を考える必要がある。SemEval で抽出対象となった情報と, 我々が抽出したい情報の比較を表 1 に示す。抽出したい情報の種類も異なるが, 我々はアスペクト表現に関しては SemEval のように名詞に限定しないことも違いである。

2.2 エンティティと属性, 評価表現の定義

SemEval で取り扱うエンティティ (E) と属性 (A) は, 小林ら [18] の意見分析抽出で用いられている 3 つ組 (対象, 属性, 評価値) のうちの “対象” と “属性” と同様である。小林らの定義では, 対象は「対象となっている商品名, サービス名」, 属性は「意見で述べられる対象の部分や性質」, 評価表現は「対象もしくは属性の質や量についての評価者の好悪を示す態度」となっている。

我々は**対象**(=エンティティ)を, これまで扱ってきた 7 種類のアスペクトのトピック (朝食, 夕食, 立地, サービス, 風呂, 部屋, 設備・アメニティ) を示す表現と定義する。朝食の場合であれば朝食を示す “朝飯”, “ブレイクファスト” などの表現, もしくは具体的なメニューである “フレンチトースト” や “クロワッサン” などである。また, **属性**と**評価表現**は小林らの定義をそのまま利用する。

2.3 アノテーションの基本方針

作業を行うデータには, 上で述べたアスペクトのトピックと評価極性が付与されてあることから, これら 2 つの情報が付与してある文を対象に, 〈対象 (E), 属性 (A), 評価表現 (P)〉の 3 つ組の同定を行う。具体例を示す。

- 朝ごはんは美味しかった。〈朝ごはん, NULL, 美味しかった〉
- 朝ごはんで食べた味噌汁の風味は最高でした。〈味噌汁, 風味, 最高〉

対象と属性は, レビュー内で明示的に書かれない場合がありその時は “NULL” とする。また, 評価表現が属性を含む場合がある。例えば “美味しい” といえば味, “高い” といえば値段といった属性を示すが, 文中には存在しないため属性は NULL のままにする。また, 着目する文中の対象または属性が NULL の場合に, レビュー内の着目文の前文に指し示す表現がある場合, ゼロ照応問題として, “NULL (実体)” という形式で記述することとする。

2.4 アノテーションの課題

実際に“朝食”に関してのレビュー文へのアノテーション作業を全体 [16] の 5%程度に対し先行して行った。前節で定義した対象、属性、評価表現は全ての文で容易に作業できる形で出現するとは限らない。作業した 789 文には肯定評価が 694 文、否定評価は 115 文、20 文に肯定否定の両方の極性が含まれていた。そのうち肯定評価文には 58 文、否定評価文には 14 文、それぞれ作業が難しい事例が含まれていた。これらの作業の難しい文を分析することでいくつかの課題に分類された。代表的な課題について事例とともに説明する。

2.4.1 属性相当を表現する節

属性は“(対象)の(属性)”という形で出現するのが典型的なパターンである。例えば“(朝食)の(地元食材)”といった形式である。しかしながら、“地元食材で作られた朝食”のような連体節で出現する場合や、上の例のような節を伴って出現する場合もある。また、節は下の例のように理由節が出現することもあるが、属性を含む場合とそうでない場合がある。アノテーションでは、属性の有無に着目して作業する必要がある。属性の種類は予め決めることができないので作業に伴い拡充することになる。

- 朝食も、地元の食材を使ってあり、美味しく頂きました。
- 朝食の量も自分に合わせて食べられるので良かったです。
- 食事は郷土料理が出てきたので、凄く良かった。

2.4.2 特殊な評価表現

評価者の態度を示す評価表現は評価極性を持つ述語で示されることが多いが、記号や数字で表現される場合がある。こうした数字や記号は述語ではないものの、評価者の態度を示すことから評価表現として扱うこととする。

また、極性を持たない名詞に存在を表す動詞“ある・いる”もしくは形容詞“ない”が続く場合に評価表現と考えられる場合があった。存在を表す動詞は、評価表現の定義である“評価者の態度”を直接表さないものの、属性の表現との組み合わせの場合は評価表現として扱う。

- 朝食はさすが!の一言です。

- しかし、料理の味は2でした。
- 特に朝食は力を入れているだけあって◎。
- 少量といいつつ、しっかりと量がありました!

2.4.3 間接評価、比較評価

ユーザの意見は必ずしも対象の属性について述べられずに、間接的な表現や他者との比較で肯定、否定を表す事例があった。これらは3つ組で捉えられないため、例外として考慮する必要がある。

- 朝食は、無料のレベルではない。
- もうけがないのではないかと心配したくらいです。
- 朝食は、あれで 1500 円以上払う気はしないでですね…。
- 朝食のバイキングは温泉旅館のものより数段上。

2.4.4 願望・提案・条件

レビュー文の中には、対象、属性、評価表現を含むものの、それらが条件付きであったり、「～はずだった」のような事実に対する意見ではない表現がある。ここではこうした文には作業を行わないこととする。

- あれもこれも食べればよかったと後悔しています。
- トマトジュースがあればもっと嬉しかったかな。
- ホントに美味しいものを少しずつ楽しみたい世代には至高の幸せです。

表 2 実験結果

手法	精度	再現率	F1
BERT _B -CRF	0.704	0.560	0.624
BERT _L -CRF	0.637	0.644	0.640
DILBERT	0.833	0.722	0.773

3 予備実験

アスペクト表現の抽出タスクを BIO チャンキングとして定式化し、評価コーパスの性質を実験によって分析した。2.3 節で説明した作業方針を踏襲し、6,012 文に対して追加のアノテーションを行い、4,761 文を学習データ、1,251 文をテストデータとして用いた。抽出手法として、固有表現抽出タスクで近年頻繁に使用されている BERT-CRF [19] を用い

表3 出力結果の例

入力系列(青色のトークンは、アスペクト表現の部分)	BERT _{large} -CRF	DILBERT
(a) 食事の蟹満喫プランは大満足でした。	{ 食事 }	{ 食事, 蟹満喫プラン }
(b) 接客が丁寧で気持ち良く、料理も夕朝共にとても美味しく、特に生きたアワビの酒蒸しなど絶品でした。	{ 接客, 料理 }	{ 接客, 料理, 生きたアワビの酒蒸し }
(c) 接客については悪くはありませんが、ホスピタリティは感じません。	{ 接客 }	{ 接客, ホスピタリティ }
(d) お料理は鮮度抜群、ブックリしたほたるいかを口にした時の幸福感!	{ お料理 }	{ お料理 }

た。また、アスペクト表現抽出の既存手法からは、コードが公開されているかつ2021年11月現在で最高値を達成したDILBERT [20]を選択した。各手法の詳細を以下に示す。

BERT-CRF アスペクトカテゴリなどのアスペクト情報は入力として与えず、トークナイズした入力系列を[CLS]トークンと[SEP]トークンで挟んだ系列をモデルの入力とした。BERTの事前学習モデルは、東北大学が提供しているBERT-base⁶⁾(以降、BERT_B-CRF)とBERT-large⁷⁾(以降、BERT_L-CRF)の2種類を用いた。

DILBERT DILBERTは、ドメイン適応のための事前学習手法であり、アスペクトカテゴリベースのマスク言語モデルタスク(以下、CMLM)とアスペクトカテゴリ予測タスクの二つからなる。標準的なマスク言語モデルは、マスクされる語がランダムに選択されるのに対して、CMLMはアスペクトカテゴリに類似した語にマスクするという特徴がある。これにより、アスペクトと思われる表現に注意が向けられることで、標準的な事前学習手法よりもアスペクト抽出タスクにおいて優れた結果を達成したという報告がある。本実験では、Nakayamaら[16]の7種類のアスペクトのトピックをDILBERTの事前学習に用いた。

表2に各手法の実験結果を示す。DILBERTがBERT-CRFとくらべて、優れたF1スコアを達成した。両手法における成功事例と失敗事例を分析することで、各手法の特長を分析するとともに本コーパスを用いた抽出タスクの難しさを調査した。表3にDILBERTとBERT_L-CRFとの出力例を示す。(a),(b),(c)は、DILBERTが全てのアスペクト表現の抽出に成功したのに対して、BERT_L-CRFは一部のアスペクト表現を取りこぼした事例である。DILBERTとBERT_L-CRF両方とも、抽出対象が「食事」や「接

客」といったレビュー中に頻出する単一名詞の表現は高確率で抽出できていた。DILBERTでは、BERT_L-CRFとくらべて「ホスピタリティ」や「接客」のような「サービス」の類義語をより獲得できる手法であることがわかった。また、DILBERTは「蟹満喫プラン」や「生きたアワビの酒蒸し」のような長い系列で表現される料理名や食事プランを抽出できるという傾向も見られた。これらは、レビュー中の語とアスペクトカテゴリ間のベクトルの類似性を事前学習時に考慮するため、アスペクトカテゴリに関連する語を積極的に抽出する傾向にあったためだと考えられる。一方(d)では、両手法とも「ぷっくりしたホタルイカ」というアスペクト表現を抽出できなかった。事例(a)と(b)にそれぞれある「蟹」や「アワビ」は学習データに含まれていたのに対して、「ホタルイカ」は学習データに含まれず対処できなかったことが原因だと考えられる。

4 おわりに

我々はこれまで、宿泊施設レビューに記述されるアスペクトカテゴリと評価極性を付与した日本語ABSAコーパスを構築してきた。本論文ではこのコーパスをもとに、更に具体的なアスペクト表現となる対象(エンティティ)とその属性および評価表現を追加付与するため、それぞれの要素の定義を行い、アノテーションの基本方針について議論と、実データが示す課題について述べた。

またBERT-CRFおよびDILBERTベースの分類器により予備分類実験を行い、その結果からそれぞれの分類器の処理傾向を考察して今後の課題を検討した。定義に従ってコーパス全体にアノテーションを行いリリースすることが今後の課題である。構築するコーパスは、国立情報学研究所の情報学研究データレポジトリ⁸⁾(IDR)にて追加公開の予定である。

6) <https://huggingface.co/cl-tohoku/bert-base-japanese-whole-word-masking>

7) <https://huggingface.co/cl-tohoku/bert-large-japanese>

8) <https://www.nii.ac.jp/dsc/idr/rakuten/>

参考文献

- [1] 新里圭司, 小山田由紀. 店舗レビューには何が書かれているか?—調査及び自動分類システムの開発—. 自然言語処理, Vol. 25, No. 1, pp. 57–79, 2018.
- [2] Qingnan Jiang, Lei Chen, Ruifeng Xu, Xiang Ao, and Min Yang. A challenge dataset and effective models for aspect-based sentiment analysis. In **Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)**, pp. 6280–6285, 2019.
- [3] Ruidan He, Wee Sun Lee, Hwee Tou Ng, and Daniel Dahlmeier. An interactive multi-task learning network for end-to-end aspect-based sentiment analysis. In **Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL)**, pp. 504–515, 2019.
- [4] Minh Hieu Phan and Philip O. Ogunbona. Modelling context and syntactical features for aspect-based sentiment analysis. In **Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL)**, pp. 3211–3220, 2020.
- [5] Jiaxin Huang, Yu Meng, Fang Guo, Heng Ji, and Jiawei Han. Weakly-supervised aspect-based sentiment analysis via joint aspect-sentiment topic embedding. In **Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)**, pp. 6989–6999, 2020.
- [6] Maria Pontiki, Dimitris Galanis, John Pavlopoulos, Harris Papageorgiou, Ion Androutsopoulos, and Suresh Manandhar. SemEval-2014 task 4: Aspect based sentiment analysis. In **Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval)**, pp. 27–35, 2014.
- [7] Maria Pontiki, Dimitrios Galanis, Haris Papageorgiou, Suresh Manandhar, and Ion Androutsopoulos. SemEval-2015 task 12: Aspect based sentiment analysis. In **Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval)**, pp. 486–495, 2015.
- [8] Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, Mohammad AL-Smadi, Mahmoud Al-Ayyoub, Yanyan Zhao, Bing Qin, Orphée De Clercq, Véronique Hoste, Marianna Apidianaki, Xavier Tannier, Natalia Loukachevitch, Evgeniy Kotelnikov, Nuria Bel, Salud María Jiménez-Zafra, and Gülşen Eryiğit. SemEval-2016 task 5: Aspect based sentiment analysis. In **Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval)**, pp. 19–30, 2016.
- [9] Yohei Seki, Lun-Wei Ku, Le Sun, Hsin-His Chen, and Noriko Kando. Overview of multilingual opinion analysis task at NTCIR-8. In **Proceedings of the Seventh NTCIR Workshop**, 2010.
- [10] 関洋平. 意見分析コーパスの現状と課題. 情報処理学会論文誌データベース (TOD), Vol. 6, No. 4, pp. 85–103, 2013.
- [11] 藤井敦, 乾孝司, 中山卓哉. レビュー解析「レビュー解析を題材とした「誤り分析マニュアル」の試作に向けた検討」. 言語処理学会第 21 回年次大会ワークショップ: 自然言語処理におけるエラー分析, 2015.
- [12] 藤井敦, 乾孝司. 自然言語処理技術の現状と展望 — エラー分析プロジェクトを通して —: 3.11 レビュー解析 — 誤り分析におけるプロセスとプロダクト—. 情報処理, Vol. 57, No. 1, pp. 30–31, 2016.
- [13] 栗原理聡, 水本智也, 乾健太郎. Twitter による評判分析を目的とした評価対象-評価表現データセット作成. 言語処理学会第 24 回年次大会発表論文集, pp. 344–347, 2018.
- [14] Haruya Suzuki, Yuto Miyauchi, Kazuki Akiyama, Tomoyuki Kajiwara, Takashi Ninomiya, Noriko Takemura, Yuta Nakashima, and Hajime Nagahara. A Japanese dataset for subjective and objective sentiment polarity classification in micro blog domain. In **Proc. of the 13th Conference on Language Resources and Evaluation (LREC 2022)**, pp. 7022–7028, 2022.
- [15] Tomoyuki Kajiwara, Chenhui Chu, Noriko Takemura, Yuta Nakashima, and Hajime Nagahara. Wrieme: A new dataset for emotional intensity estimation with subjective and objective annotations. In **Proc. of 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies**, pp. 2095–2104, 2021.
- [16] **A Large-Scale Japanese Dataset for Aspect-based Sentiment Analysis**, 2022.
- [17] 中山祐輝, 村上浩司, Ikuko Hardaway. アスペクトベース意見分析における日本語評価コーパスの構築. 言語処理学会第 27 回年次大会発表論文集, pp. 375–380, 2021.
- [18] 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. 3, pp. 203–222, 2005.
- [19] Fábio Souza, Rodrigo Nogueira, and Roberto Lotufo. Portuguese named entity recognition using bert-crf. <https://arxiv.org/abs/1909.10649>, 2019.
- [20] Antony Lekhtman, Yftah Ziser, and Roi Reichart. DIL-BERT: Customized pre-training for domain adaptation with category shift, with an application to aspect extraction. In **Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing**, pp. 219–230. Association for Computational Linguistics, Nov 2021.