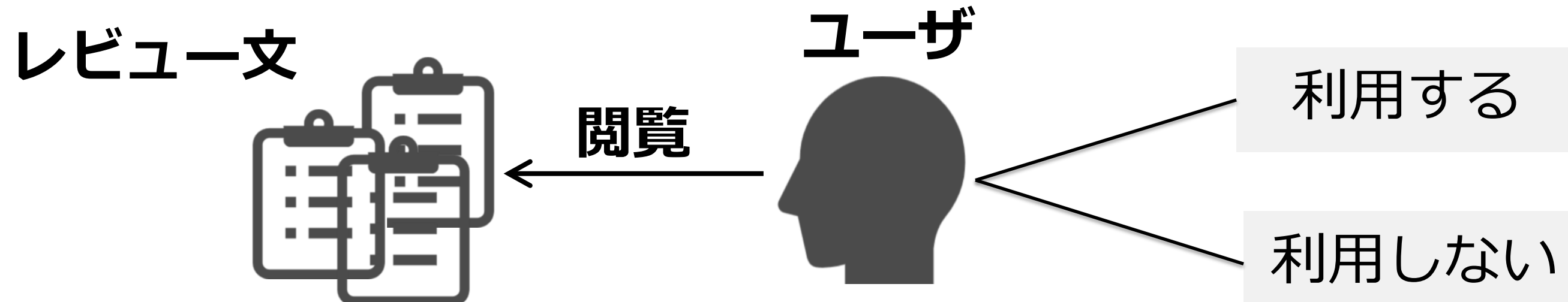




1. はじめに

◆レビューの重要性と問題点

- 消費者の意思決定に影響 [1]



- 膨大なレビュー全てに目を通すのは困難 [2]

必要な情報のみを抽出して**閲覧負担を軽減**することが必要

◆閲覧負担の軽減

- 評価に関する**属性と属性値**がレビューの要約に有用 [4]
- 評価に関する**属性と属性値**を抽出して**評価ラベルに分類**することで要約可能 [5][6]

例) サービスがよかったです。
属性 : 商品のある側面
属性値 : 属性に対する評価

属性-属性値ペアの分類のためにペア抽出が必要

[1] Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Suresh Manandhar and Ion Androutsopoulos. (2015). "SemEval-2015 Task 12: Aspect Based Sentiment Analysis", Association for Computational Linguistics, *Proceedings of the 9th International Workshop on Semantic Evaluation*, pp. 486-495.
[2] Ziyu Guan, Long Chen, Wei Zhao, Yi Zheng, Shulong Tan and Deng Cai. (2016). "Weakly-Supervised Deep Learning for Customer Review Sentiment Classification", *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, pp. 3719-3725.
[3] Yequan Wang, Minlie Huang, Xiaoyan Zhu and Li Zhao. (2016). "Attention-based LSTM for Aspect-level Sentiment Classification", Association for Computational Linguistics, *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pp. 606-615.
[4] Muhammad Bilal, Huma Israr, Muhammad Shahid and Amin Khan. (2016). "Sentiment classification of Roman-Urdu opinions using Naive Bayesian, Decision Tree and KNN classification techniques", *Journal of King Saud University - Computer and Information Sciences*, Vol.28, No.3, pp. 330-344.
[5] Hongliang Dai and Yangqiu Song. (2019). "Neural Aspect and Opinion Term Extraction with Mined Rules as Weak Supervision", Association for Computational Linguistics, *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp.5268-5277.

2. 先行研究

◆英語レビューに対する既存の抽出ルール例 [5]

- 属性が主語で属性値が述語 (形容詞)
例) The service was good.
依存関係 : **主語名詞**
- 属性が主語で属性値 (動詞)が述語
例) Cleaning is completed
依存関係 : **主語名詞**
- 属性が属性値 (動詞)の目的語
例) I enjoyed cuisine.
依存関係 : **目的語**

[5] Hongliang Dai and Yangqiu Song. (2019). "Neural Aspect and Opinion Term Extraction with Mined Rules as Weak Supervision", Association for Computational Linguistics, *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp.5268-5277.

3. 提案手法

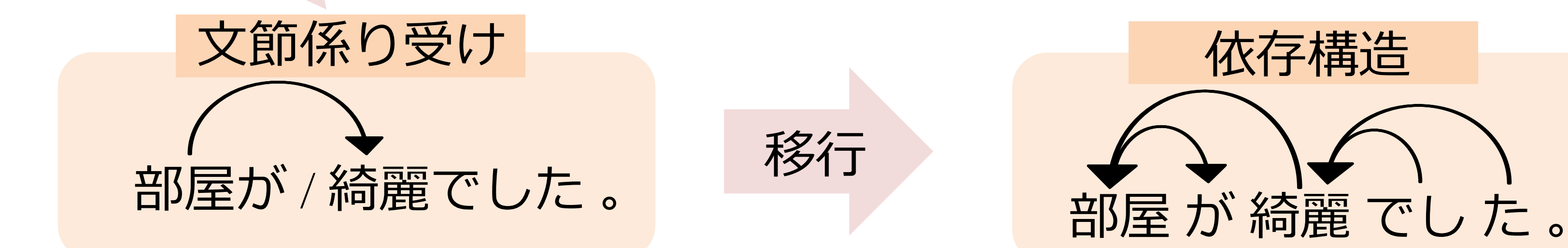
属性-属性値ペアの抽出

新規性

日本語レビューから依存構造を用いて属性-属性値ペアを抽出

◆属性-属性値ペア抽出の背景

- Universal Dependencies (UD) [6]に基づき, 形態素の依存構造に徐々に移行
→ 言語間で共通化した構文構造アノテーション
→ 依存構造の使用で言語資源作成の省力化が可能 [7]



◆属性-属性値ペア抽出の抽出方法

属性値 : 評価表現辞書 [8] から, 属性値となりうる形容詞・動詞・名詞 [4] の語を抽出
属性 : UD に基づいた自然言語処理ライブラリGiNZA [9] を使用し属性値と特定の依存関係にある語を抽出

◆日本語の特徴と適用する際の問題点

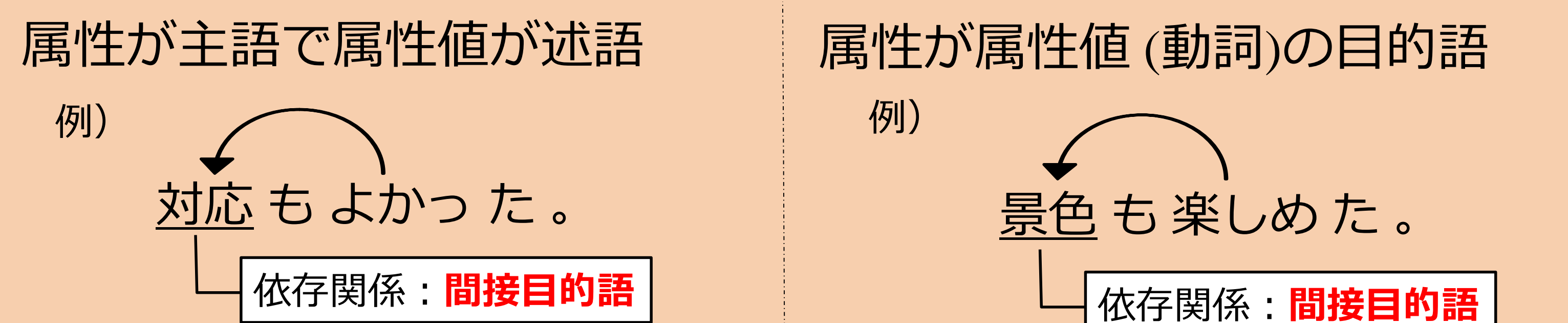
- 「が」がつく名詞に依存関係ラベル『主語名詞』を付与 [10]
『主語名詞』 : 主語である名詞に付与される依存関係ラベル

主語につく助詞が「が」以外の時依存関係ラベルが『間接目的語』になる

- 「を」がつく名詞に「目的語」と付与 [10]
『目的語』 : 目的語である名詞に着けられる依存関係ラベル

目的語につく助詞が「が」以外の時依存関係ラベルが『間接目的語』になる

◆日本語レビューに適用するため追加した提案ルール



[6] Universal Dependencies contributors. "Universal Dependencies", <https://universaldependencies.org/>, 最終閲覧日 : 2020/11/23
[7] Hiroshi Kanayama and Koichi Takeda (2017). "Multilingualization of Question Answering Using Universal Dependencies", *Proceedings of OKBQA 2017*, pp.13-16
[8] 小林のぞみ. 「evaluative expressions」, http://www.syncha.org/evaluative_expressions.html, 最終閲覧日 : 2020/11/12
[9] 松田寛, 大村舞, 浅原正幸. (2019). 「短単位品詞の用法曖昧性解決と依存関係ラベリングの同時学習」, 言語処理学会第 25 回年次大会講演論文集, F2-3.
[10] 浅原正幸, 金山博, 宮尾祐介, 田中貴秋, 大村舞, 村脇有吾, 松本裕治. (2019) 「Universal Dependencies 日本語コーパス」, 自然言語処理, Vol.26, No.1, pp.2-26

4. 評価実験

表1 : 使用データ数

| | 褒め | 苦情 | 評価なし | 中立 | 要求 | マルチラベル | 合計 |
|-------|-------|-----|------|-----|-----|--------|-------|
| 使用データ | 1,979 | 740 | 950 | 101 | 230 | 309 | 4,309 |

◆使用データ

- TSUKUBAコーパス[11]
- 評価ラベルが付与された楽天トラベルのレビュー文

◆ペア抽出精度の比較

正解データ : 人手で抽出した全てのペア

| 表2 : 抽出精度 | 再現率 | 適合率 | F値 |
|-----------|--------------|--------------|--------------|
| 英語圏ルール | 0.143 | 0.273 | 0.188 |
| 提案手法 | 0.265 | 0.319 | 0.289 |

レビュー文単位に精度を測り平均をとる
正解のないレビュー文は評価に含めない

◆分類モデルの設定

正解データ : 提案手法で抽出できたペアに対して人手で感情ラベルを付与

- レビュー内容を属性-属性値ペアで表現できるBag of Words を特徴量に
- 短文で不均衡なデータに有効な Negation Naive Bayes [12] を使用
- 要求文は「評価」では無いため属性-属性値ペアで表現できない
→ ルールベース [11] で排除

◆分類結果

表3 : 分類精度

| | 褒め | 苦情 | 評価なし | 中立 | 全体 |
|-----------|--------------|--------------|--------------|--------------|--------------|
| 英語圏ルール正解率 | 0.856 | 0.737 | 0.527 | 0.000 | 0.824 |
| 英語圏ルールF値 | 0.888 | 0.705 | 0.338 | 0.000 | 0.515 |
| 提案手法正解率 | 0.877 | 0.756 | 0.692 | 0.450 | 0.852 |
| 提案手法F値 | 0.902 | 0.718 | 0.374 | 0.503 | 0.624 |

提案手法により

- 英語圏ルールで抽出できないペアも抽出可能になり, 抽出精度が向上
- 属性-属性値のペアの分類が可能であるか検証 → 分類可能**

[11] Rakuten Institute of Technology 「楽天データ公開」, https://rit.rakuten.co.jp/data_release_ja, 最終閲覧日 : 2020/11/12
[12] 古宮 嘉那子, 伊藤 裕佑, 佐藤 直人, 小谷 善行. (2013). 「文書分類のための Negation Naive Bayes」, 自然言語処理, Vol.20, No.2, pp. 161-182.
[13] 大森 晃. (2019). 「要求文同定論 - 要求とは何か? その文は要求を表現しているか?」, 情報処理学会論文誌, Vol.60, No.3, pp.976-1023.

5. おわりに

- 英語圏の属性-属性値抽出ルールを日本語レビューに適用・ルールを拡張
- 日本語レビューから依存構造を用いて属性-属性値ペアを抽出可能
- レビューを要約する評価ラベルごとに分類
- 英語圏ルール使用時より高い精度で評価ラベルに分類が可能

→ **日本語レビューから属性-属性値ペアを抽出し評価の分類を実現**

◆今後の課題

- データセット内に存在する複数ラベルが振られたデータにも対応する