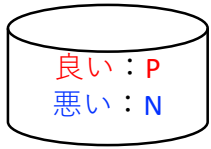


●研究背景

・印象極性辞書

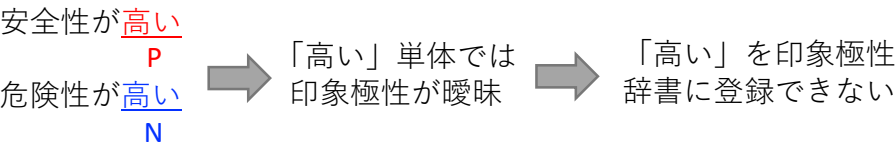
日本語評価極性辞書(用言編)



- ・評価値表現辞書[1][2]内の5,280件の用言とその印象極性が収録されている
- ・人手で用言に印象極性を付与しているため、収録内容が正確

・印象極性辞書の問題点

- ・極性が曖昧な語句である、極性曖昧性[3]を持つ語句が収録されていない



・極性曖昧性を持つ語句を印象極性辞書へ登録する研究[4]

- ・文脈一貫性による印象極性推定[5]を利用する
- ・<値段, 高い>のような、<属性, 評価表現>で表される**評価情報**[6]の形にし、極性曖昧性を緩和させた状態で印象極性辞書へ登録する

画質は**良かったけど**、**値段**が高い。
P 逆接 ?

画質は**良かったけど**、**値段**が高い。
P 逆接 N

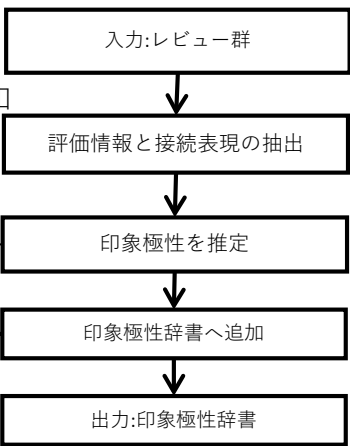
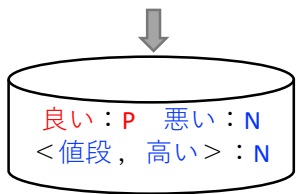


図1: 印象極性辞書の作成フロー[4]

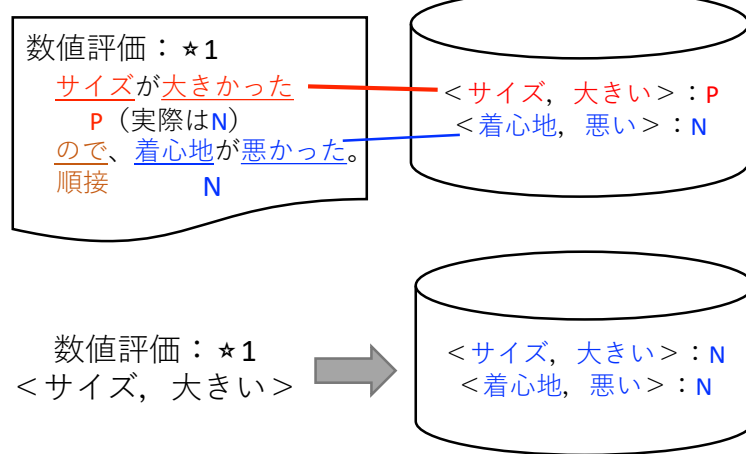
[課題]: 辞書に収録される印象極性の信頼性が低い

極性曖昧性を持つ語句を印象極性辞書へ登録する研究[4]を経て登録語の数は増やすことができたが、付与された印象極性の信頼性は低い。

その理由は以下の2つである。

- ・<サイズ, 大きい>など、評価情報の形にしても印象極性が曖昧なものがある
- ・機械での印象極性の付与は、人手での付与よりも精度が低い

[目的]: レビューに付与された**数値評価**を考慮して印象極性の推定を行うことで、辞書の信頼性を高める



●データ調査

- ・レビュー中の接続表現よりも前方の部分、後方の部分それぞれと、数値評価の関係を人手で調査した。

- ・楽天トラベルデータ[7]の750件のレビューを使用した

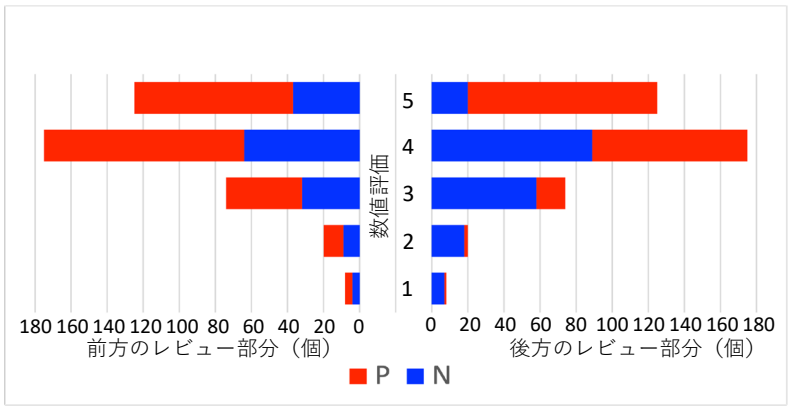


図2: 接続表現の前後方部分と数値評価の関係

- 図2, より,
- ・後方部分と数値評価に関係性がある
 - ・後方部分において、数値評価が1〜3の時は**N**, 5の時は**P**と仮定できる

●提案手法

- ・従来手法[4]の図1のフローに、図3中の4番目の処理である数値評価での印象極性修正処理を追加する。

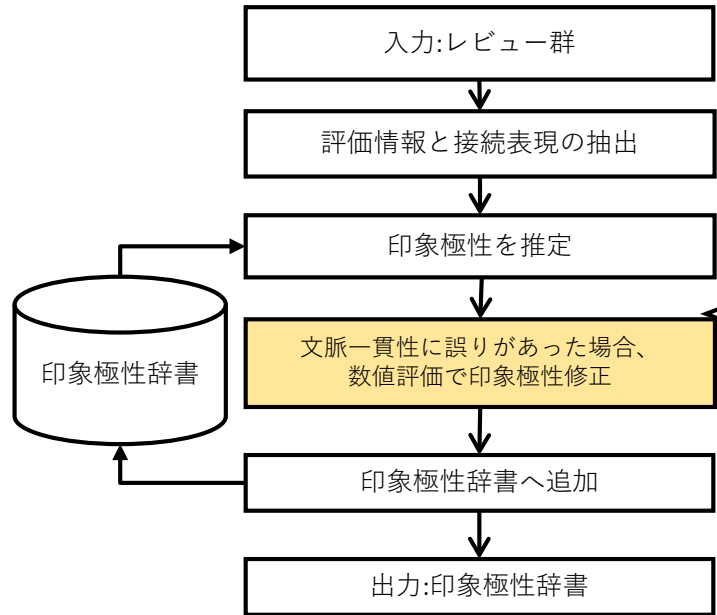


図3: システム全体の処理の流れ

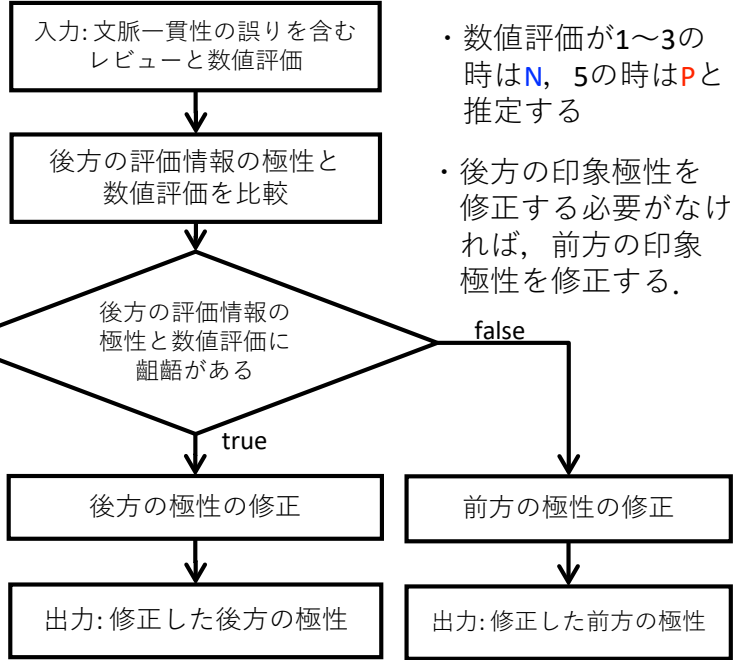


図4: 図3中の印象極性修正処理の詳細

●実験結果

- ・楽天トラベルデータ[7]の10,000件のレビューを使用して提案手法を実施し、2,216個の評価情報が新たに印象極性辞書に登録された。
- ・辞書の登録語を増やす手法[4]（比較手法）と、提案手法の印象極性の推定精度を比較した。（結果：表1）
- ・新たに登録された評価情報を、接続表現の前方に出てきたものと後方に出てきたものに分け、各数値評価との関係性を調査した。（結果：図5）

表1: 比較手法と提案手法の極性推定精度

	Pの推定精度	Nの推定精度
従来手法	0.59	0.49
提案手法	0.64	0.53

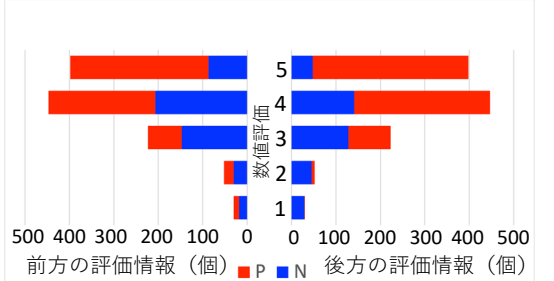


図5: 接続表現の前後方の評価情報と数値評価の関係

●結論

実験結果より、比較手法と比べて提案手法の**P**の語句、**N**の語句それぞれの推定精度が向上し、辞書の信頼性が高まった。また、図5より、仮定通り数値評価が1,2の時には**N**が多く、5の時には**P**が多くなったことから、数値評価を極性推定に利用することは有効であることが分かった。しかし、数値評価が3の時に**P**と**N**の両方が多く現れたため、3の場合の推定に誤りが多く含まれたことが分かった。

●謝辞

本研究では、国立情報学研究所のIDRデータセット提供サービスにより楽天グループ株式会社からご提供いただいた「楽天データセット」(https://rit.rakuten.com/data_release/)を利用しました。心より感謝申し上げます。

●参考文献

[1] 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.
[2] 小林のぞみ, 乾 健太郎, 松本 裕治. 意見情報の抽出/構造化のタスク仕様に関する考察. 情報処理学会研究報告 NL171-18, pp.111-118, 2006.01.
[3] 吉原 圭祐, 中山 碧土, 早坂 絵央, 王 羽, 片倉 多智, 木村 優平, 西村 百之輔, 真鍋 大雅, 藤田 和成, 延澤 志保, “印象極性辞書のための語句の持つ極性曖昧性の推定,” NLP 若手の会第 15 回シンポジウム, 2020.
[4] 水崎 佑迅, 大和 矢 悠仁, 阿部 遥佳, 延澤 志保, “印象極性辞書作成のための評価情報の極性推定,” NII-IDR ユーザ フォーラム, no.P13, 2021.
[5] 那須川 哲哉, 金山 博, “文脈一貫性を利用した極性付評価表現の語彙獲得,” 情報処理学会研究報告, vol.2004-NL-162, pp.109–116, 2004.
[6] 平山 拓央, 湯本 高行, 新居 学, 高橋 豊, “属性評価モデルに基づく商品評価の抽出と提示,” DEIM Forum 2011
[7] 楽天グループ株式会社 (2020): 楽天トラベルデータ. 国立情報学研究所情報学研究データリポジトリ. (データセット). <https://doi.org/10.32130/idr.2.2>