

研究背景

研究目的

・ 極性辞書による極性推定

・ 極性が曖昧な語句の辞書登録

日本語評価極性辞書

極性辞書の問題点

→ 極性が曖昧な語句の存在

例) 危険性が高い 安全性が高い

属性と評価表現のペアにすることで
曖昧性を減らし、極性付与可能[2]

<危険性,高い> <安全性,高い>

属性, 評価表現 (評価情報) を辞書の見出しとし、辞書の構築を行うことで曖昧な語句も辞書への追加を可能にし、問題点の改善を行った[3].

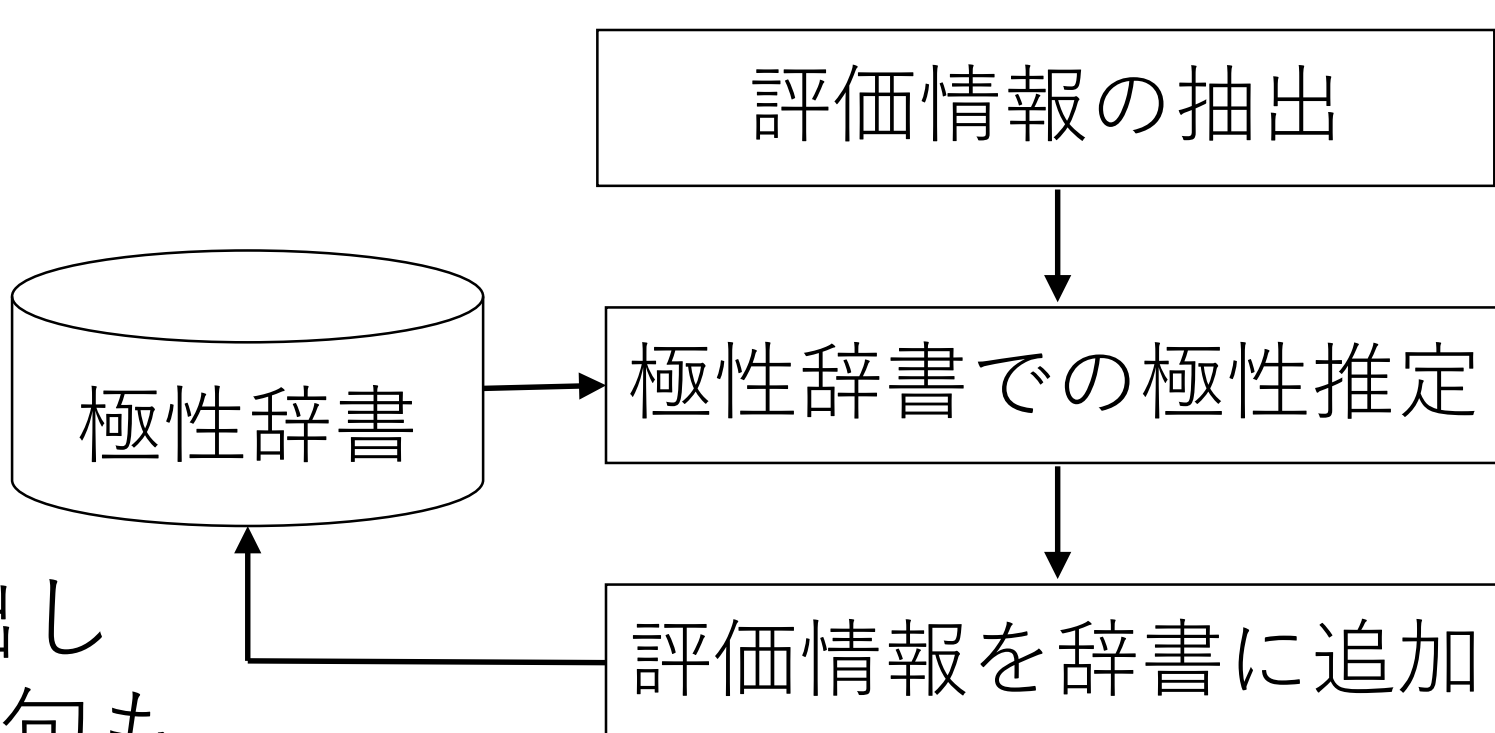


図1：辞書構築の流れ[3]

・ 研究課題

評価情報を使用しても極性が曖昧なものがある.

例) <値段,安い>, <サイズ,大きい> など

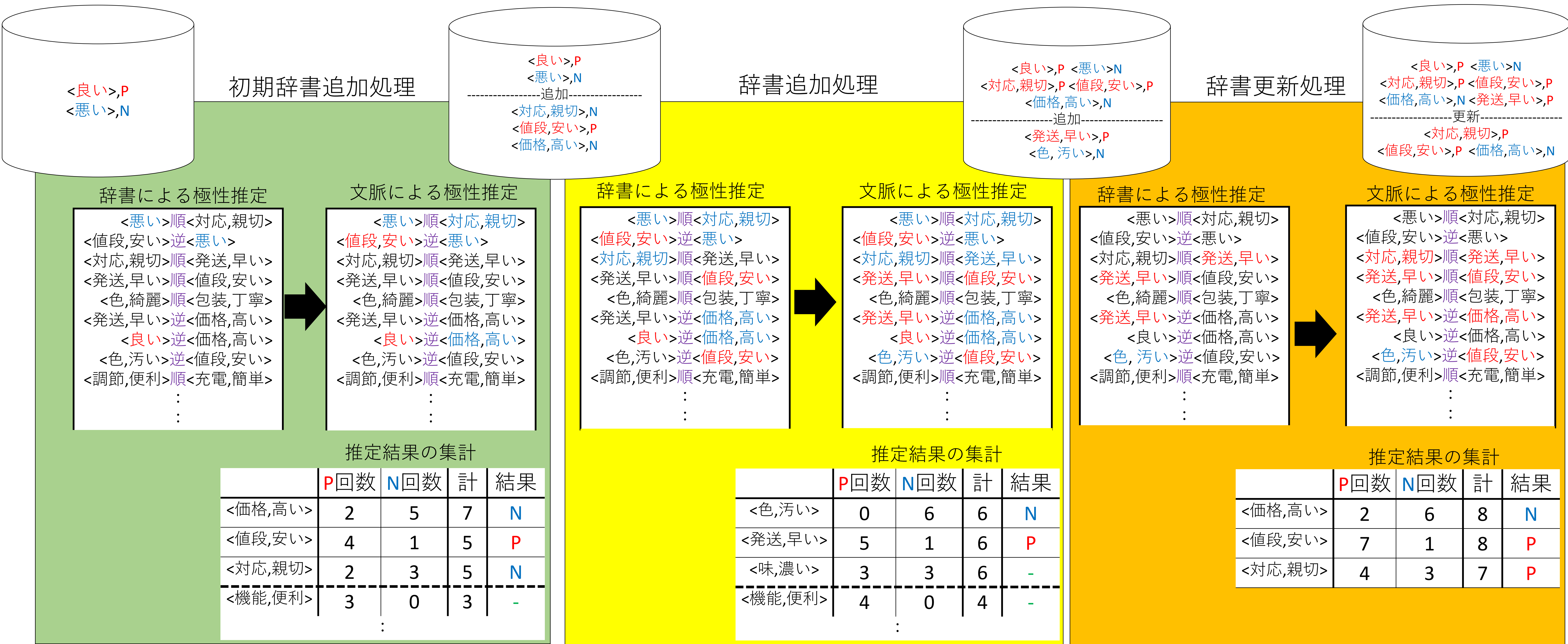
また、本来の極性とは異なる誤った推定を行うと辞書に追加される極性も誤ってしまう。
→ 辞書の信頼性が問題になる.

評価情報を見出しとした極性辞書の構築を行う.

評価情報に対して辞書への追加と更新処理の2段階の推定をすることで信頼性の高い辞書の構築を行う.

提案手法

実験結果



極性推定，極性更新の結果

- ・ 楽天市場レビューデータ2019年12月 10,000レビューを使用.
- ・ 表1に提案手法と正解データの比較結果，表2に極性辞書の更新を行わない辞書拡張の手法（比較手法）と本手法と極性に違いが出るものの比較を行った.

U:手法のうち，極性を見つけれられなかったもの,E:正解データのうち，極性判定ができないもの

表1：提案手法と正解データとの比較

		提案手法		
		P	N	合計
正解データ	P	67	0	67
	N	7	7	14
	合計	74	7	81

表3：両手法と正解データの比較

	更新なし	提案手法
正解率	0.88	0.91

表2：辞書の極性が異なった評価情報の比較

評価情報		比較手法			提案手法		
		P	N	U	P	N	U
<量,多い>	P	2	1	4	7	0	0
	N	1	2	0	3	0	0
	E	0	0	0	0	0	0
<サイズ,ない>	P	2	2	0	3	1	0
	N	0	1	0	0	0	1
	E	0	0	0	0	0	0
<ワン,大きい>	P	2	1	1	4	0	0
	N	0	0	0	0	0	0
	E	0	2	0	1	0	1

結論

- ・ 表1, 表3から提案手法の正解率が91%となり，辞書の更新処理を行わない比較手法よりも高い精度を示していることから信頼性の高い極性辞書の構築を行うことができた.

謝辞

参考文献

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様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。心より感謝いたします。

- [1] 小林 ひとみ, 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情報の収集,” 自然言語処理, Vol.12, No.3, pp203-222, 2005.
- [2] 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.
- [3] 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,” IDRユーザーフォーラム 2021.
- [4] 楽天グループ株式会社 (2020): 楽天市場データ. 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.2.1>