

WRIME:主観と客観の感情極性分類 のための日本語データセット

鈴木陽也[†] 宮内裕人[†] 秋山和輝[†] 梶原智之[†] 二宮崇[†]
武村紀子[‡] 中島悠太[‡] 長原一[‡]

愛媛大学[†] 大阪大学[‡]

<https://github.com/ids-cv/wrime>

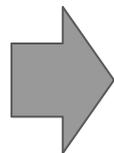


背景

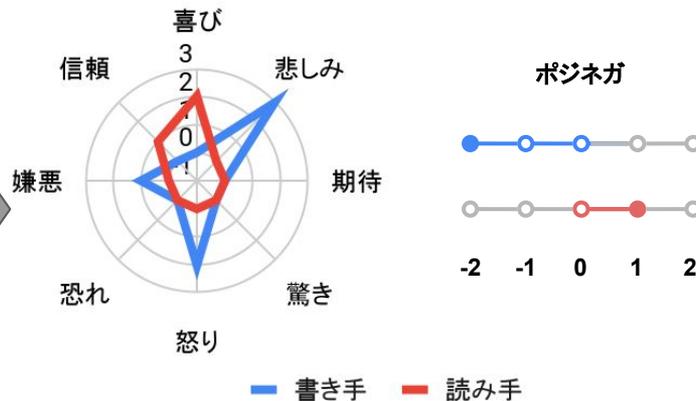
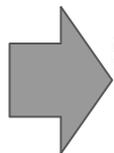
感情分析:

テキストから人間の感情強度や感情極性を予測するタスク

ダルゴナコーヒーおいち！



感情分析
モデル



課題:

感情分析における日本語データセットの不足

提案

- 主観・客観、感情強度・感情極性のすべてをラベル付けした日本語コーパスを構築

ラベル: Plutchikの基本8感情の強度(0:無、1:弱、2:中、3:強)

ポジネガ(-2、-1、0、1、2)

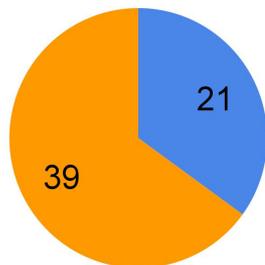
	感情極性	感情強度	主観	客観	データ件数
筑波大学文単位評価極性タグ付きコーパス	✓	×	×	✓	4,309
Twitter日本語評判分析データセット	✓	×	×	✓	534,962
WRIME	✓	✓	✓	✓	35,000

主観ラベルのアノテーション

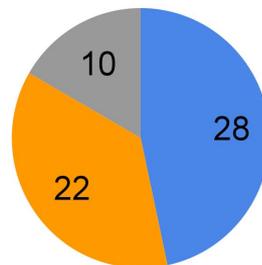
- 60人のアノテータをクラウドソーシングで雇用
- アノテータは自身のSNS投稿データに対し、主観的な感情をラベル付け

クラウドソーシング: Lancers (<https://www.lancers.jp/>)

データ: 2010年から2020年のSNS投稿データ



● 男性 ● 女性



● 20代 ● 30代 ● 40代～

客観ラベルのアノテーション

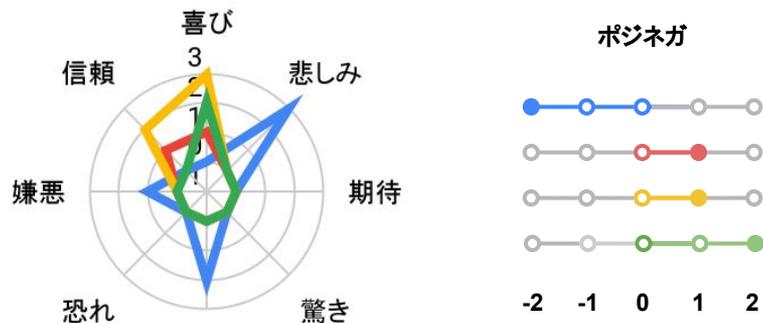
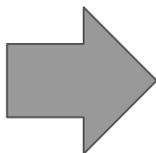
- 3人のアノテータをクラウドソーシングで雇用
- 全投稿に対し、客観的な感情(書き手の感情を推測)をラベル付け

クラウドソーシング: Lancers (<https://www.lancers.jp/>)

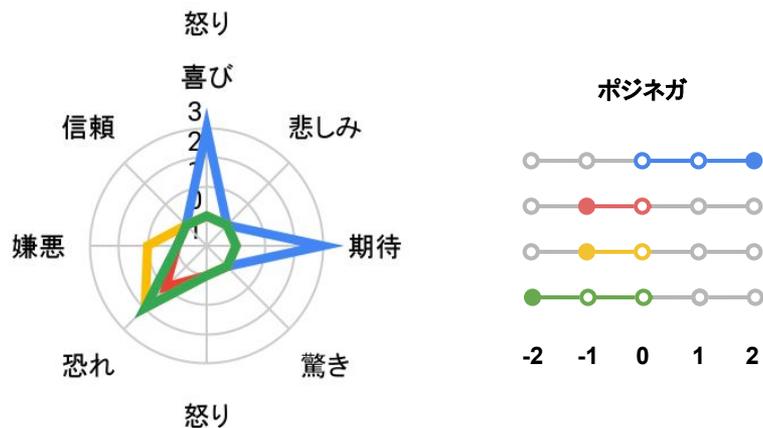
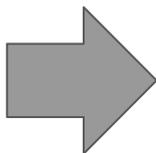
アノテータ: 30代の女性が2人、40代の女性が1人

アノテーションの例

ダルゴナコーヒーおいち！



電車で今鬼緊張してる..



— 書き手 — 読み手A — 読み手B — 読み手C

コーパス使用例

- 感情ごとにモデルを構築
- 順序分類タスクとしてQuadratic Weighted Kappaで評価
- データセット:
訓練用(40人30,000件)、検証用(10人2,500件)、評価用(10人2,500件)

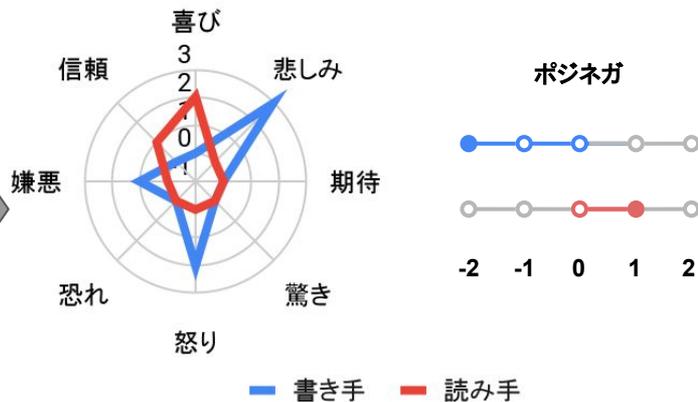
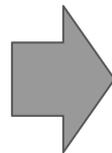
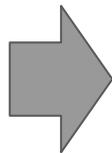
	喜び	悲しみ	期待	驚き	怒り	恐れ	嫌悪	信頼	ポジネガ
BoW	0.420	0.218	0.190	0.219	0.180	0.174	0.208	0.049	0.348
BERT	0.539	0.381	0.308	0.337	0.264	0.195	0.234	0.135	0.512

WRIME: 主観と客観の感情強度分類のための 日本語データセット

課題: 感情分析における日本語コーパスの不足

提案: 主観・客観、感情強度・感情極性をラベル付けした日本語コーパスの提案

ダルゴナコーヒーおいち！



文献情報

Ver.1: 基本感情のみ

- 梶原智之. [WRIME: 主観と客観の感情強度を付与した日本語データセット](#). 自然言語処理, Vol.28, No.3, pp.907-912, 2021.
- Tomoyuki Kajiwara, Chenhui Chu, Noriko Takemura, Yuta Nakashima, Hajime Nagahara. [WRIME: A New Dataset for Emotional Intensity Estimation with Subjective and Objective Annotations](#). In Proceedings of the 2021 Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), pp.2095-2104, 2021.
- 梶原智之, Chenhui Chu, 武村紀子, 中島悠太, 長原一. [主観感情と客観感情の強度推定のための日本語データセット](#). 言語処理学会第27回年次大会, pp.523-527, 2021.

Ver.2: 基本感情＋感情極性

- 宮内裕人, 鈴木陽也, 秋山和輝, 梶原智之, 二宮崇, 武村紀子, 中島悠太, 長原一. [主観と客観の感情極性分類のための日本語データセット](#). 言語処理学会第28回年次大会, pp.1495-1499, 2022.
- <https://github.com/ids-cv/wrime>