



## データセットの構築

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

- ・ SNS上の日本語テキスト35,000件
- ・ Plutchikの8感情 (4段階強度)  
+ 感情極性 (5段階強度) を付与
- ・ 主観 (書き手) と客観 (読み手)  
両方をアノテーション

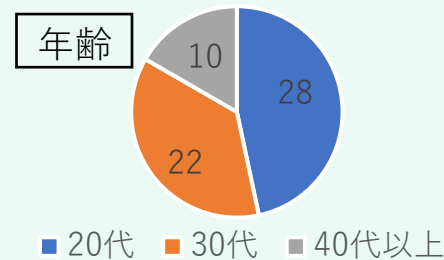
傘盗まれたんだけど！！

	喜び	悲しみ	期待	驚き	怒り	恐れ	嫌悪	信頼	ポジネガ
主観	0	2	0	0	2	0	3	0	-2
客観A	0	0	0	0	2	0	0	0	-1
客観B	0	3	0	0	3	0	3	0	-2
客観C	0	0	0	0	3	0	0	0	-2

## アノテーション詳細

## 主観アノテーション

- ・ ランサーズで60人雇用  
(男性21名、女性39名)
- ・ 自分の投稿にラベル付け



## 客観アノテーション

- ・ ランサーズで3人雇用  
(30代女性2、40代女性1)
- ・ 35,000件全てにラベル付け

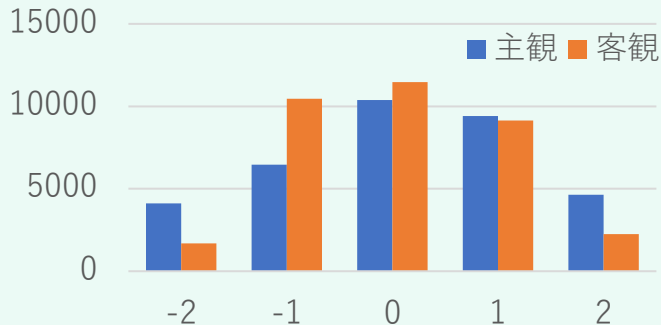
アノテーション一致率

QWK (極性)

客観-客観	0.530~0.688
主観-客観	0.493~0.605
主観-客観平均	0.621

## データセットの分析

(件数)



- ・ 客観主観の比較

客観→2,-2が少なく-1が多い

- ・ 基本感情と極性の関係性

	喜び	期待	信頼	驚き	悲しみ	恐れ	怒り	嫌悪
主観ポジネガ	0.59	0.38	0.30	0.05	-0.53	-0.30	-0.35	-0.47
客観ポジネガ	0.66	0.40	0.25	0.04	-0.54	-0.41	-0.23	-0.47

## BERTによる感情極性推定

Quadratic Weighted Kappa

	主観	客観
BoW+LogReg	0.348	0.408
東北大BERT	0.512	0.695
hottoSNS-BERT(主観)	0.558	0.627
hottoSNS-BERT(客観)	<b>0.595</b>	<b>0.743</b>

- ・ 訓練：30,000 検証：2,500 評価：2,500
- ・ 主観の推定の方が困難