

WRIME: A New Dataset for Emotional Intensity Estimation with Subjective and Objective Annotations

Tomoyuki Kajiwara^{†*} Chenhui Chu^{‡*}

Noriko Takemura* Yuta Nakashima* Hajime Nagahara*

[†] Graduate School of Science and Engineering, Ehime University, Japan

[‡] Graduate School of Informatics, Kyoto University, Japan

* Institute for Datability Science, Osaka University, Japan



Emotional Intensity Estimation

We work on the task of estimating the intensity of **emotions** such as *joy* and *sadness* from the user-generated text.

Previous works used the following basic emotions.

- **Russell's emotions**

valence, arousal, and dominance

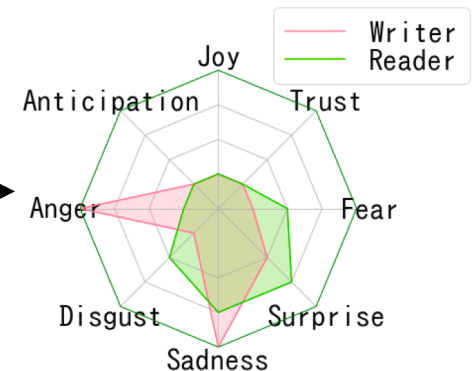
- **Ekman's emotions**

anger, disgust, fear, joy, sadness, and surprise

- **Plutchik's emotions**

anger, disgust, fear, joy, sadness, surprise, trust, and anticipation

The tire of my car was flat.
I heard that it might be mischief.



Emotional Intensity Datasets

	Emotion	Intensity	Annotator	Lang.	Size
ISEAR	Ekman	n/a	Writer	En	7,666
Blogs	Ekman	{Low, Med., High}	Reader	En	5,025
SemEval-2007	Ekman	[0, 100]	Reader	En	1,250
WASSA-2017	Joy, Sadness,	[0, 1]	Reader	En	7,097
SemEval-2018	Anger, and Fear	[0, 1]	Reader	En	12,634
EmoBank	Russell	{1, 2, 3, 4, 5}	Reader	En	10,062
GoodNewsEveryone	Plutchik	{Low, Med., High}	Reader	En	5,000
WRIME	Plutchik	{No, Low, Med., High}	Both	Ja	17,000

Emotional Intensity Annotation in Previous Works

- Readers estimate the writers' emotions.

Emotional Intensity Annotation in Our Work

- Writers label their own emotions.
- Readers also estimate and label the writers' emotions.

→ **We can analyze the difference between *subjective (writers')* and *objective (readers')* emotions.**

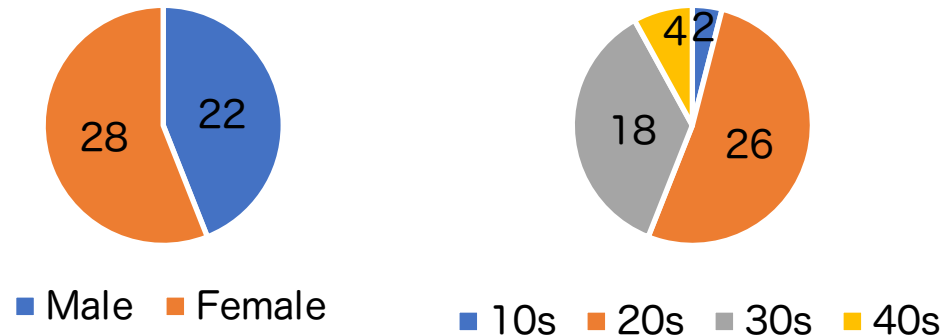
Table of Contents

- Introduction
- **Emotional Intensity Annotation**
- Analysis of Our Dataset
- Experiments on Emotional Intensity Estimation

Annotating Subjective Labels

We hired 50 participants via crowdsourcing service.

- *Lancers* <https://www.lancers.jp/>
- They copy and paste their own past SNS posts.
- And then labeled the posts with the *subjective* emotions.
 - Plutchik's eight emotional intensity
 - Four-point scale (0: No, 1: Weak, 2: Medium, 3: Strong)
 - Each participants labeled 100 to 500 posts



Our Dataset

- 17,000 Japanese posts from 50 participants
- 9-year range from June 2011 to May 2020

Annotating Objective Labels

We hired 3 annotators via crowdsourcing service.

- *Lancers* <https://www.lancers.jp/>
- They labeled all the posts with the *objective* emotions.
 - Two women in their 30s and one woman in their 40s.

An expanded version is available. →
(40,000 posts from 80 participants)



The tire of my car was flat.
I heard that it might be mischief.

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

	Joy	Sadness	Anticipation	Surprise	Anger	Fear	Disgust	Trust
Writer	0	3	0	1	3	0	0	0
Reader 1	0	3	0	3	1	2	1	0
Reader 2	0	2	0	2	0	0	0	0
Reader 3	0	2	0	2	0	1	1	0

Table of Contents

- Introduction
- Emotional Intensity Annotation
- **Analysis of Our Dataset**
- Experiments on Emotional Intensity Estimation

Distribution of Emotion Intensity

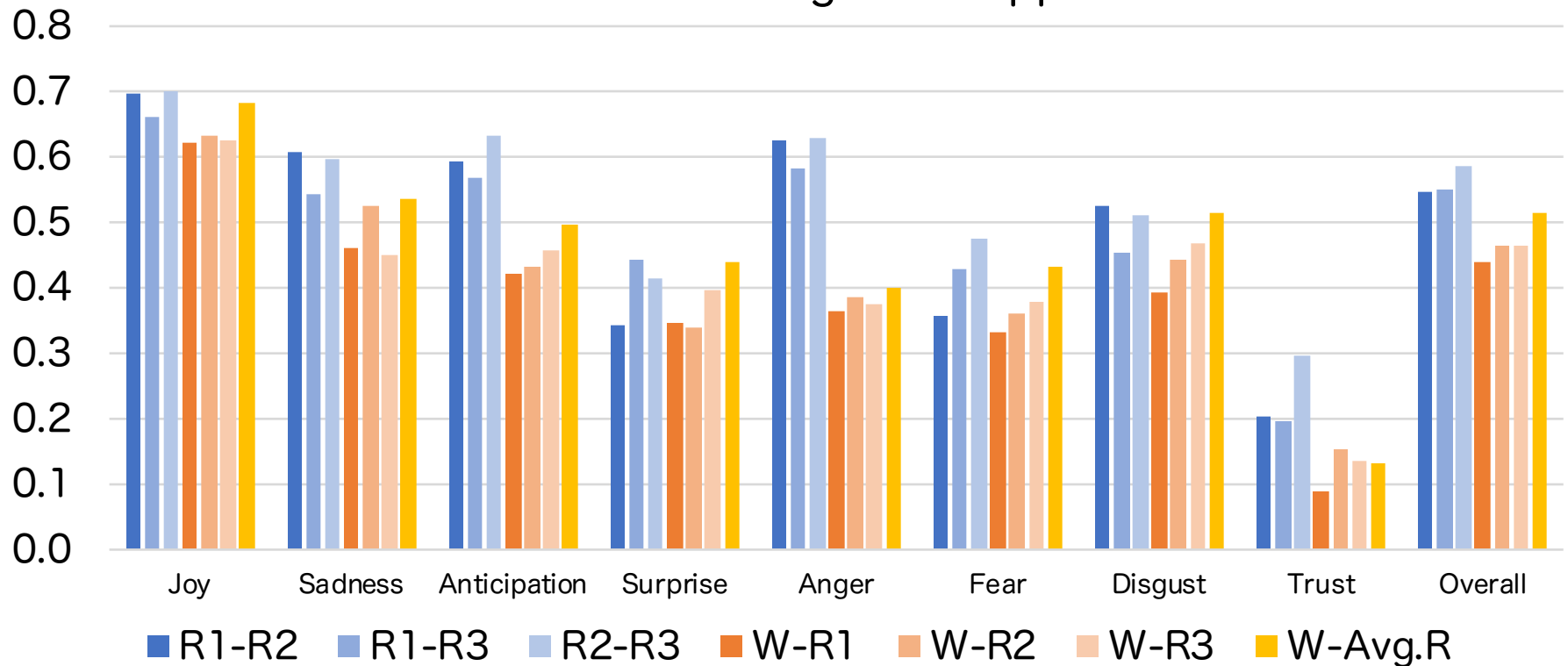
- For all emotions, intensity 0 is most frequently assigned.
- This is not surprising, as it is rare for a single post to come with many emotions, which may be contradictory to each other, at the same time.
- However, for emotions of *anger* and *trust*, readers may tend to underestimate the emotions of the writers. 95% of labels have an intensity 0.

Intensity	Joy				Sadness				Anticipation			
	W	R1	R2	R3	W	R1	R2	R3	W	R1	R2	R3
0	9,942	12,043	12,379	11,213	10,472	13,205	11,961	12,559	9,991	11,714	12,509	10,796
1	2,454	291	1,074	1,397	2,837	389	1,881	2,123	2,996	610	1,245	2,683
2	2,283	2,285	2,055	3,475	2,140	2,127	2,168	1,846	2,172	2,507	1,825	2,119
3	2,321	2,381	1,492	915	1,551	1,279	990	472	1,841	2,169	1,421	1,402
Intensity	Surprise				Anger				Fear			
	W	R1	R2	R3	W	R1	R2	R3	W	R1	R2	R3
0	11,148	10,534	14,143	10,974	14,408	16,278	16,180	16,223	13,355	13,285	15,163	13,877
1	2,605	997	1,429	2,840	1,284	156	304	311	1,815	384	626	1,478
2	1,778	3,234	971	2,027	661	285	315	266	1,082	2,032	838	1,070
3	1,469	2,235	457	1,159	647	281	201	200	748	1,299	373	575
Intensity	Disgust				Trust				Overall			
	W	R1	R2	R3	W	R1	R2	R3	W	R1	R2	R3
0	13,258	14,538	14,333	11,959	12,682	16,165	15,920	15,979	95,256	107,762	112,588	103,580
1	1,882	449	1,248	2,436	2,162	470	466	609	18,035	3,746	8,273	13,877
2	959	1,190	934	1,242	1,239	239	395	300	12,314	13,899	9,501	12,345
3	901	823	485	1,363	917	126	219	112	10,395	10,593	5,638	6,198

Difference between Writers & Readers (1/2)

- **Writer-Reader** agreement is lower than that of **Reader-Reader**.
- **Anger**: There is a large gap between WR-agreement and RR-agreement.
- **Trust**: Both WR-agreement and RR-agreement are low.

Quadratic Weighted Kappa



Difference between Writers & Readers (2/2)

Overall, readers tend to underestimate the writers' emotions.

- Readers overlook about 40% of the writers' strong emotions.
- Especially in emotions of *anger* and *trust*, readers overlook more than 60% of the writers' strong emotions.

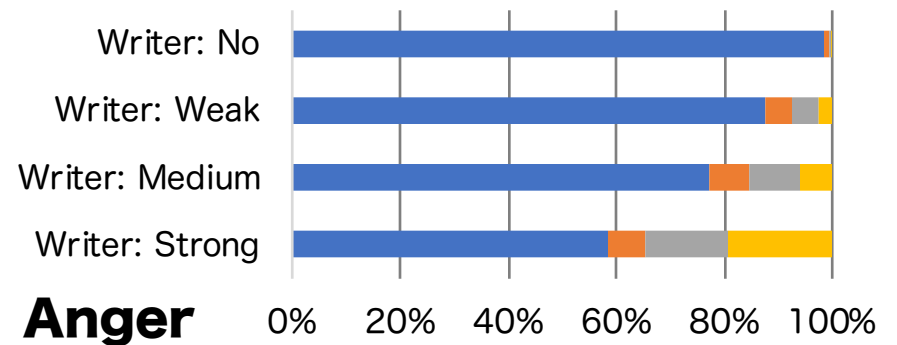
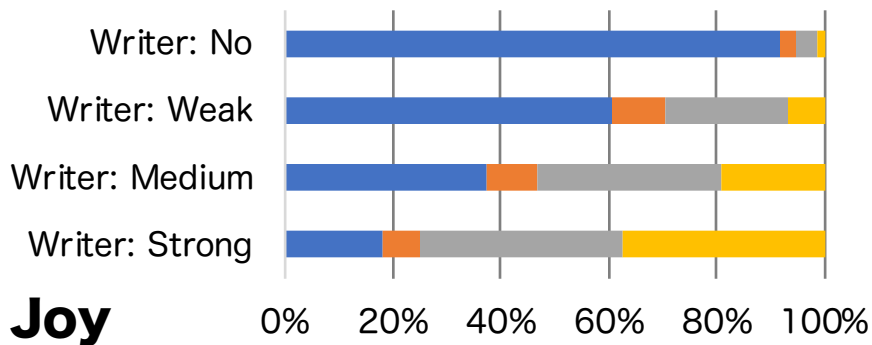
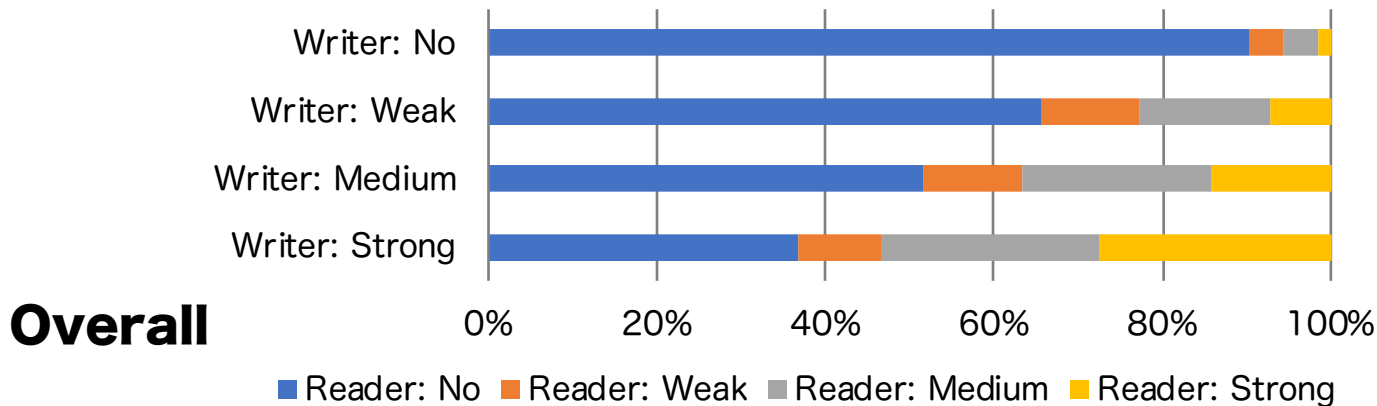


Table of Contents

- Introduction
- Emotional Intensity Annotation
- Analysis of Our Dataset
- **Experiments on Emotional Intensity Estimation**

Estimating Writers' Subjective Emotional Intensity

- Train: 15,000 Dev: 1,000 Test: 1,000 Metric: MAE
- Modal Class: A baseline that always outputs an intensity 0.
- Obj. BERT: A model trained by the readers' labels.
- Subj. BERT: A model trained by the writers' labels.
- Surprisingly, Obj. BERT achieved the best performance.
- For *anger* and *trust*, all models always output an intensity 0.

	Joy	Sadness	Anticipation	Surprise	Anger	Fear	Disgust	Trust	Overall
Modal Class	0.90	0.71	0.91	0.68	0.22	0.34	0.44	0.43	0.578
Obj. BERT	0.67	0.62	0.79	0.63	0.22	0.36	0.43	0.43	0.519
Subj. BERT	0.73	0.67	0.90	0.68	0.22	0.34	0.44	0.43	0.553
Med. Readers	0.53	0.53	0.72	0.69	0.20	0.52	0.52	0.43	0.522

Estimating Readers' Objective Emotional Intensity

- Train: 15,000 Dev: 1,000 Test: 1,000 Metric: MAE
- Modal Class: A baseline that always outputs an intensity 0.
- Obj. BERT: A model trained by the readers' labels.
- Subj. BERT: A model trained by the writers' labels.
- Estimating the readers' labels is easier than that of writers.
- Obj. BERT has also achieved the highest performance.

	Joy	Sadness	Anticipation	Surprise	Anger	Fear	Disgust	Trust	Overall
Modal Class	0.60	0.46	0.71	0.52	0.04	0.42	0.38	0.03	0.395
Obj. BERT	0.40	0.41	0.48	0.44	0.04	0.39	0.35	0.02	0.317
Subj. BERT	0.49	0.45	0.69	0.52	0.04	0.41	0.38	0.04	0.377
Med. Readers	0.22	0.25	0.28	0.35	0.05	0.28	0.27	0.03	0.215

WRIME: A New Dataset for Emotional Intensity Estimation



- We annotated 17,000 Japanese SNS posts from 50 crowdsourced workers with emotional intensity.
- Unlike existing datasets in which readers estimated the writers' emotions, this study collected the writers' own emotional intensity.
- We found that readers tend to underestimate the writers' emotions.
- Even the strong emotions of writers cannot be detected by the readers, especially in the emotions of **anger** and **trust**.
- Experimental results on emotional intensity estimation show that it is more difficult to estimate the writers' *subjective* emotions than the readers' *objective* ones.

The tire of my car was flat.
I heard that it might be mischief.

