Emotional Intensity Estimation based on Writer's Personality

Haruya Suzuki[†]   **Sora Tarumoto**[†]   **Tomoyuki Kajiwara**[†]
Takashi Ninomiya[†]   **Yuta Nakashima**[‡]   **Hajime Nagahara**[‡]
[†]Ehime University   [‡]Osaka University
{suzuki@ai.,tarumoto@ai.,kajiwara@,ninomiya@}cs.ehime-u.ac.jp
{n-yuta,nagahara}@ids.osaka-u.ac.jp

Abstract

We propose a method for personalized emotional intensity estimation based on a writer's personality test for Japanese text. Existing emotion analysis models are difficult to accurately estimate the writer's subjective emotions behind the text. We personalize the emotion analysis using not only the text but also the writer's personality information. Experimental results show that personality information improves the performance of emotional intensity estimation. Furthermore, a hybrid model combining the existing personalized method with ours achieved state-of-the-art performance.

1 Introduction

Emotional intensity estimation (Strapparava and Mihalcea, 2007; Bostan et al., 2020; Kajiwara et al., 2021) is one of the major challenges in the natural language processing community with many applications in dialogue systems (Tokuhisa et al., 2008) and social media mining (Stieglitz and Dang-Xuan, 2013). Emotional intensity estimation predicts the (often discretized) intensities of finer-grained emotions, such as Ekman's basic emotions, i.e., *joy*, *sadness*, *surprise*, *anger*, *fear*, and *disgust* (Ekman, 1992) and Plutchik's basic emotions, i.e., *joy*, *sadness*, *expectation*, *surprise*, *anger*, *fear*, *disgust*, and *trust* (Plutchik, 1980).

WRIME[1] (Kajiwara et al., 2021; Suzuki et al., 2022) is a corpus from Social Networking Service (SNS) text in Japanese for emotional intensity estimation. As exemplified in Table 1, the corpus adopts Plutchik's basic emotions from *both* the writers' (*subjective*) and the readers' (*objective*) points of view. Their experimental results showed that estimating subjective emotion is more difficult than objective emotion. This fact renders an additional challenge to subjective emotional intensity estimation. That is, there can be a latent factor that
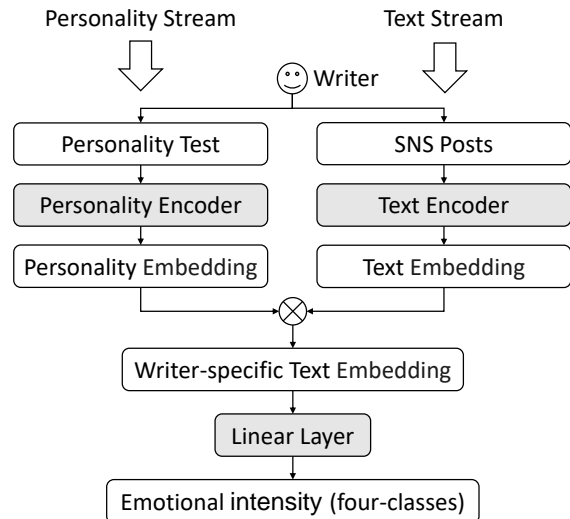
---

[1] https://github.com/ids-cv/wrime

Figure 1: An overview of the proposed method.

modulates the superficial emotion perceived from the text *per se*.

A straightforward hypothesis to explain the difference is that *the writer's personality affects their writing*. This hypothesis seems plausible as the same text can have different meanings depending on who wrote it, the contexts such as the writer's preceding SNS text and the circumstance the writer is in, etc; the writer's personality can influence all these aspects and can alter how they author text.

This hypothesis inspires us to design a model specialized for subjective emotion. The model uses the personality test result of each writer, which is fortunately included in the corpus. Specifically, given the personality test result, which is answers to 60 questions (Saito et al., 2001) based on the Big Five personality five-factor model (Goldberg, 1992), we embed 60 answers into a high-dimensional feature vector. Our model, shown in Figure 1 combines feature vectors from the SNS text and the personality to improve the estimation performance.

	Joy	Sadness	Anticipation	Surprise	Anger	Fear	Disgust	Trust

I often have expectations for people I meet in real life, not on the web,
and am a little disappointed due to overly high expectations. What should I do?

	Joy	Sadness	Anticipation	Surprise	Anger	Fear	Disgust	Trust
Subjective	0	3	0	0	0	0	0	0
Objective A	0	1	0	0	0	0	1	0
Objective B	0	1	2	0	0	0	0	0
Objective C	0	2	0	0	0	0	0	0
BERT	0	1	1	0	0	0	3	0
+ Personality	0	3	0	0	0	0	0	0

Why can't people who work hard be rewarded for their efforts? It's so frustrating.

	Joy	Sadness	Anticipation	Surprise	Anger	Fear	Disgust	Trust
Subjective	0	3	0	0	2	0	1	0
Objective A	0	2	0	0	0	2	2	0
Objective B	0	2	0	0	0	0	3	0
Objective C	0	3	0	0	0	0	0	0
BERT	0	3	0	0	0	0	0	0
+ Personality	0	3	0	0	2	0	3	0

Table 1: The upper rows of each table show examples of emotional intensity labels, consisting of subjective and objective ones, where three annotators (A–C) were invited for this sample (0: none, 1: weak, 2: medium, 3: strong). The lower part of each table shows the prediction results of the baseline model and our method.

Experimental results on the WRIME corpus show that our model performs better than both Bag-of-Words (BoW) and BERT (Devlin et al., 2019) baselines without personality information, which suggests the advantage of using writers' personality information for subjective emotional intensity estimation. Furthermore, a hybrid model combining the existing personalized method (Milkowski et al., 2021) with ours achieved state-of-the-art performance in emotional intensity estimation. The performance is on par with the performance of our human annotators.

2 Related Work

Human emotions are subjective and have personal biases depending on many factors such as the first language, age, education (Wich et al., 2020; Al Kuwatly et al., 2020), gender (Bolukbasi et al., 2016; Tatman, 2017), race (Sap et al., 2019; Davidson et al., 2019), and personality (Kajiwara et al., 2021). Due to the nature of such personal biases, writers may express different emotions even if they wrote the same text (Milkowski et al., 2021; Ngo et al., 2022). Taking into account the emotional differences between writers is important for a high-quality emotional analysis.

Personalized emotion analysis has been studied in recent years. Milkowski et al. (2021) personalized the emotion analysis by focusing on the labeling variation among annotators. They proposed Personal Emotional Bias (PEB) as a measure of labeling variation and showed that such user-specific information contributes to emotional intensity estimation. Kajiwara et al. (2021) personalized the emotion analysis by focusing on the personality of the text writer. They considered personality information based on the Big Five personality five-factor model (Goldberg, 1992) in a simple way (concatenation or attention) and showed that such user-specific information contributes to emotional intensity estimation. This study advances the latter approach and proposes a more effective method to model personality information for this task.

3 Methods

As argued in Section 1, we hypothesize that the writer's personality influences how they express themselves. We thus propose to leverage the personality of the writer as auxiliary information, especially for subjective emotional intensity estimation.

Figure 1 shows the overall structure of our model, which consists of the *text stream* and *personality*

stream, fused together for estimating personality-aware emotional intensities. The text stream is the feature extractor of a basic emotional intensity estimation model, and the personality stream can also be seen as the feature extractor of regression model, trained to predict individual personality traits of the Big Five taxonomy (Goldberg, 1992).

3.1 Text Stream

Our text stream is a part of simple BERT pre-trained model (Devlin et al., 2019)-based emotional intensity classifiers. The 768-dimensional feature vector h_t corresponding to `[CLS]` token is fed into a linear classifier for each emotion to predict one of the four-level intensities of the emotion (as in Table 1). We use h_t as text features.

3.2 Personality Stream

The WRIME (Kajiwara et al., 2021; Suzuki et al., 2022) corpus provides a personality assessment result for each writer during the curation process. This personality assessment is based on the Big Five model (Goldberg, 1992), and our writers were asked to answer 60 questions related to talkative-ness, anxiousness, etc. (Saito et al., 2001) over a seven-point scale. The answers are collectively mapped into continuous likeliness values (Big Five Scales) of the writer having the five personality traits (i.e. *extraversion*, *neuroticism*, *openness*, *conscientiousness*, and *agreeableness*).

For embedding a writer's personality in a feature vector, we mimic the process of computing the likeliness values from the 60 answers using a 3-layer multilayer perceptron with a 60-dimensional input layer and a 5-dimensional output layer, as shown in Figure 2. The middle layer's dimensionality is 768, which is the same as the output of BERT. We use the middle layer as personality feature h_p.

3.3 Fusion of Text and Personality Streams

The feature vectors h_t and h_p are fused for personality-aware emotional intensity estimation, where the dimensionalities of the feature vectors are both $d = 768$. We exploratively evaluate the following four approaches for fusion.

1. **Difference** uses $h_{\text{diff}} = |h_t - h_p|$ as a fused vector. This approach retains the dimensionality of the fused vector h.

2. **Product** applies the element-wise multiplication $h_{\text{prod}} = h_t \odot h_p$. This approach retains the dimensionality of the fused vector h.
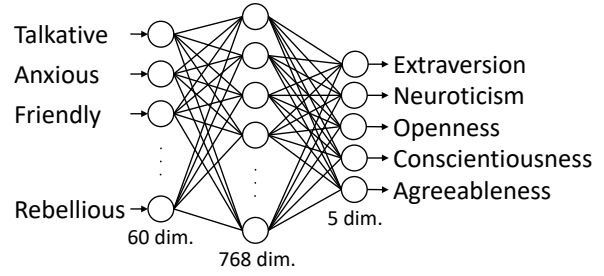

Figure 2: Mapping from 60 answers to Big Five personality traits.

3. **Concatenation** is given by $h_{\text{conc}} = [h_t, h_p]$, where $[\cdot, \cdot]$ is the operator for concatenation. This approach doubles the fused vector's dimensionality.

4. **All** concatenate all these fused vectors, i.e., $h_{\text{all}} = [h_{\text{diff}}, h_{\text{prod}}, h_{\text{conc}}]$. This approach results in a $4d$-dimensional fused vector.

For fusion approach $f \in \{\text{diff}, \text{prod}, \text{conc}, \text{all}\}$, emotional intensity is estimated by

$$y_e = \text{softmax}(W_e h_f + b_e), \quad (1)$$

where $y_e \in [0,1]^4$ is the confidences of four intensity levels for emotion e in Plutchik's basic emotions (Plutchik, 1980), and $W_e \in \mathbb{R}^{4 \times D_f}$ and $b_e \in \mathbb{R}^4$ are parameters of the classifier for emotion e (D_f is size of fused vector for approach f).

4 Experiments

Using WRIME (Kajiwara et al., 2021; Suzuki et al., 2022), a corpus for estimating the emotion analysis in Japanese, we conduct an experiment to evaluate a four-class (i.e. none, weak, medium, and strong) classification of writers' emotional intensity.

4.1 Setting

4.1.1 Dataset

For a fair comparison with the previous work (Milkowski et al., 2021), we first split 35,000 SNS posts by 60 writers into two parts: One is for training/evaluating the models, while the other is for computing the user representation in PEB. Following Milkowski et al. (2021), the latter part thus contains past 15% of SNS posts authored by each writer. The former is further split into training, validation, and evaluation sets. The training, validation, and evaluation sets respectively contain 25,500 posts from 40 writers, 2,125 posts from 10 writers, and 2,125

	Joy	Sadness	Anticipation	Surprise	Anger	Fear	Disgust	Trust	Overall
BoW	0.307	0.181	0.151	0.132	0.165	0.145	0.178	0.080	0.227
+ Difference	0.313	0.206	0.164	0.144	0.151	0.117	0.168	0.108	0.229
+ Product	0.293	0.233	0.139	0.145	0.164	0.154	0.200	0.037	0.231
+ Concat	0.294	0.217	0.148	0.120	0.144	0.145	0.188	0.101	0.236
+ All	0.300	0.231	0.169	0.111	**0.167**	0.108	0.178	0.097	0.230
+ Pc	0.293	0.193	0.153	0.121	0.135	0.153	0.151	0.066	0.219
+ Pa	0.310	0.192	0.130	0.121	0.138	0.093	0.180	0.067	0.213
+ PEB	0.329	0.292	**0.207**	0.198	0.147	0.174	0.181	**0.142**	0.260
+ Personality (All)	**0.336**	**0.312**	0.199	**0.200**	0.147	**0.185**	**0.249**	0.115	**0.281**
BERT	0.551	0.419	0.352	0.341	0.375	0.302	0.431	0.206	0.437
+ Difference	0.559	0.444	0.368	0.336	0.381	0.313	0.410	0.225	0.440
+ Product	0.573	0.468	0.363	**0.351**	0.384	0.311	**0.439**	0.240	0.459
+ Concat	0.558	0.453	0.332	0.331	0.359	0.303	0.433	0.222	0.444
+ All	0.573	**0.476**	0.373	0.345	0.404	**0.328**	0.425	0.153	0.454
+ Pc	0.564	0.443	0.377	0.310	0.358	0.290	0.403	**0.243**	0.438
+ Pa	0.560	0.430	0.359	0.322	0.392	0.284	0.413	0.206	0.429
+ PEB	0.576	0.455	0.377	0.336	**0.421**	0.327	0.429	0.198	0.451
+ Personality (All)	**0.588**	0.469	**0.389**	0.343	0.394	0.311	0.451	0.214	**0.462**
Annotator 1	0.622	0.461	0.423	0.348	0.363	0.333	0.394	0.089	0.439
Annotator 2	0.633	0.526	0.432	0.339	0.386	0.361	0.442	0.153	0.465
Annotator 3	0.624	0.450	0.459	0.396	0.374	0.380	0.467	0.134	0.463

Table 2: Quadratic weighted kappa of the writer's subjective emotional intensity estimation.

posts from 10 writers. We employ quadratic weighted kappa[2] (Cohen, 1968) as our evaluation metric, which assesses the agreement between the estimated and correct labels, considering the ordinal nature of our labels.

4.1.2 Implementation Details

For the text steam, we evaluated the two models.

- **BoW** extracts bag-of-words from a post and estimates emotional intensity by linear regression model. MeCab (IPADIC-2.7.0)[3] (Kudo et al., 2004) is used for word segmentation.

- **BERT** is a Japanese BERT[4] (Devlin et al., 2019) with a structure of 12 layers, 12 attention heads, and 768 dimensions, pre-trained with mask language modeling objectives on 86 million Japanese Twitter posts.

The BoW model is implemented using scikit-learn[5] (Pedregosa et al., 2011). The HuggingFace

Transformers (Wolf et al., 2020) is used to implement the BERT model. BERT is fine-tuned using the cross entropy loss with the batch size of 32 posts and the dropout rate of 0.1. The learning rate is set to 2e-5 with the Adam optimizer (Kingma and Ba, 2015). Early stopping is used for training and training stops when the metric (quadratic weighted kappa) of the validation set does not improve for 3 epochs. For linear regressor of the BoW model is trained with the learning rate of 0.01.

Both BoW and BERT models are coupled with writers' personality features in Section 3.2 by the four fusion approaches. For this personality embedding, the multilayer perceptron shown in Figure 2 with sigmoid activation is trained for 1,000 epochs with the SGD optimizer and the mean squared error loss.

4.1.3 Comparative Methods

We compare the following three existing methods with the proposed method.

- **Pc** (Kajiwara et al., 2021) uses $h_c = W_c[u, v]$ as a feature vector, where v is a 768-dimensional textual representation corresponding to the [CLS] token of BERT and u is a 786-dimensional personality representation computed by a linear mapping from the

5-dimensional Big Five personality traits. We use Equation (1) as classifier with replacing h_f with h_c.

- **Pa (Kajiwara et al., 2021)** employs the scaled dot-product attention (Vaswani et al., 2017) as $h_a = \text{attention}(\boldsymbol{W}^Q \boldsymbol{u}, \boldsymbol{W}^K \boldsymbol{v}, \boldsymbol{W}^V \boldsymbol{v})$ for feature extraction, so that textual representation corresponding to the `[CLS]` token of BERT can be weighted based on the writer's personality. Emotional intensity estimation is done in the same way as Pc but with h_a.

- **PEB (Milkowski et al., 2021)** extracts features by $h_{\text{PEB}} = \boldsymbol{W}_{\text{PEB}}[\boldsymbol{z}, \boldsymbol{v}']$, where $\boldsymbol{v}'$ is a textual representation given by linearly transforming $\boldsymbol{v}$ into a 50-dimensional vector and $\boldsymbol{z}$ is a user representation given by linear transformation of a 8-dimensional vector representing annotation bias for each emotion into a 50-dimensional vector. Again, Equation (1) is used with h_{PEB} for emotional estimation.

4.2 Results

Table 2 shows the experimental results. The scores are the average of quadratic weighted kappa values over three training runs, where we trained the models five times with different parameter initialization and excluded the maximum and minimum kappa values. The table is divided into three blocks: The top two are for the emotional intensity estimation models, while the bottom block shows the human performance of three annotators in the WRIME corpus (Kajiwara et al., 2021; Suzuki et al., 2022). Note that these annotators do not know the writer's personality or past posts.

Compared to the BoW model, the BERT model consistently achieves higher performance. This is a reasonable result for two reasons: feature extraction with BoW cannot take context into account, and BoW does not have the benefit of a large-scale corpus such as the one used for pre-training BERT.

The proposed methods showed improvement in many emotions compared to the baseline model, which does not take the writer's personality into account. Our Difference method improved performance on five out of eight emotions for BoW and on six emotions for BERT. Our Product method improved performance on half of the eight emotions for BoW and consistently improved performance on all emotions for BERT. While our Concat method only improved performance on three out

of eight emotions for BoW, it improved on five emotions for BERT. Our All method improved performance on half of the eight emotions for BoW and on six emotions for BERT. Furthermore, the proposed methods consistently improved performance in the overall evaluation. These experimental results confirm the effectiveness of the proposed methods for estimating subjective emotional intensity with the writer's personality information.

Next, we discuss the results of a comparison of the proposed and existing methods. The existing methods for Pc and Pa (Kajiwara et al., 2021) did not show significant improvement from each baseline model in the overall evaluation in this experimental setting. Although these existing methods utilize the writer's personality similar to our method, they differ in the method for feature extraction from the personality information. In contrast, our methods consistently improved performance in the overall evaluation.

Another existing method, PEB (Milkowski et al., 2021), achieves higher performance than our methods for BoW and comparable performance to our methods for BERT in the overall evaluation. Because our method, which takes into account the personality of the writer, and PEB, which takes into account labeling variations, take different approaches to personalize emotional intensity estimation, we can expect synergies from their combination. The bottom methods in Table 2, using $h_{\text{hybrid}} = [h_{\text{diff}}, h_{\text{prod}}, h_{\text{conc}}, h_{\text{PEB}}]$ instead of h_f, achieved the best performance for both BoW and BERT models in the overall evaluation. Furthermore, BERT with both writer's personality and PEB achieved performance comparable to the human annotators in the overall evaluation. These experimental results demonstrate the usefulness of personality information in emotional intensity estimation and the effectiveness of our feature extraction method from the personality test.

The bottom row of each table in Table 1 shows examples of output from our model. By taking into account the personality of the writer, we succeeded in emphasizing the emotional intensity of *sadness* in the upper example and *anger* in the lower example, respectively. In the personality test, these writers answered strongly to the questions "pessimistic" and "irascible," respectively.

5

5 Conclusions

To improve the performance of estimating subjective emotional intensity by writers, we propose an emotional intensity estimation model that takes into account the writer's personality information. In the proposed method, we first extracted feature representations from the results of a personality test based on the Big Five personality five-factor model. Then, we fused that personality features with textual features from BoW or BERT to personalize the emotional intensity estimation. Experimental results on subjective emotional intensity estimation in Japanese SNS text reveal the effectiveness of the proposed methods in taking into account the personality of the writer.

Currently, our method requires writers to answer a 60-item personality test. Therefore, our future work includes studying methods for estimating the writer's personality from their past posts, and how to combine them with the present method.

Acknowledgements

References

Hala Al Kuwatly, Maximilian Wich, and Georg Groh. 2020. Identifying and Measuring Annotator Bias Based on Annotators' Demographic Characteristics. In *Proceedings of the Fourth Workshop on Online Abuse and Harms*, pages 184–190.

Tolga Bolukbasi, Kai-Wei Chang, James Zou, Venkatesh Saligrama, and Adam Kalai. 2016. Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings. In *Proceedings of the 30th Conference on Neural Information Processing Systems*, pages 4356–4364.

Laura Ana Maria Bostan, Evgeny Kim, and Roman Klinger. 2020. GoodNewsEveryone: A Corpus of News Headlines Annotated with Emotions, Semantic Roles, and Reader Perception. In *In Proceedings of the 12th Language Resources and Evaluation Conference*, pages 1554–1566.

Jacob Cohen. 1968. Weighted Kappa: Nominal Scale Agreement Provision for Scaled Disagreement or Partial Credit. *Psychological Bulletin*, 70(4):213–220.

Thomas Davidson, Debasmita Bhattacharya, and Ingmar Weber. 2019. Racial Bias in Hate Speech and Abusive Language Detection Datasets. In *Proceedings of the Third Workshop on Abusive Language Online*, pages 25–35.

Jacob Devlin, Ming-Wei Chang, Kenon Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4171–4186.

Paul Ekman. 1992. A General Psychoevolutionary Theory of Emotion. *Cognition and Emotion*, 6(3–4):169–200.

Lewis R Goldberg. 1992. The Development of Markers for the Big-Five Factor Structure. *Psychological Assessment*, 4(1):26–42.

Tomoyuki Kajiwara, Chenhui Chu, Noriko Takemura, Yuta Nakashima, and Hajime Nagahara. 2021. WRIME: A New Dataset for Emotional Intensity Estimation with Subjective and Objective Annotations. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 2095–2104.

Diederik P. Kingma and Jimmy Lei Ba. 2015. Adam: A Method for Stochastic Optimization. In *Proceedings of the 3rd International Conference on Learning Representations*.

Taku Kudo, Kaoru Yamamoto, and Yuji Matsumoto. 2004. Applying Conditional Random Fields to Japanese Morphological Analysis. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, pages 230–237.

Piotr Milkowski, Marcin Gruza, Kamil Kanclerz, Przemyslaw Kazienko, Damian Grimling, and Jan Kocon. 2021. Personal Bias in Prediction of Emotions Elicited by Textual Opinions. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing: Student Research Workshop*, pages 248–259.

Anh Ngo, Agri Candri, Teddy Ferdinan, Jan Kocon, and Wojciech Korczynski. 2022. StudEmo: A Non-aggregated Review Dataset for Personalized Emotion Recognition. In *Proceedings of the 1st Workshop on Perspectivist Approaches to NLP*, pages 46–55.

Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12(85):2825–2830.

Robert Plutchik. 1980. A General Psychoevolutionary Theory of Emotion. *Theories of Emotion*, 1:3–31.

Takako Saito, Tomoyasu Nakamura, Toshihiko Endo, and Madoka Yokoyama. 2001. Standardization of Big Five Scales Using the Adjective Check List. *Kyushu University Psychological Research*, 2:135–144.

Maarten Sap, Dallas Card, Saadia Gabriel, Yejin Choi, and Noah A. Smith. 2019. The Risk of Racial Bias in Hate Speech Detection. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1668–1678.

Stefan Stieglitz and Linh Dang-Xuan. 2013. Emotions and Information Diffusion in Social Media — Sentiment of Microblogs and Sharing Behavior. *Journal of Management Information Systems*, 29(4):217–248.

Carlo Strapparava and Rada Mihalcea. 2007. SemEval-2007 Task 14: Affective Text. In *Proceedings of the Fourth International Workshop on Semantic Evaluations*, pages 70–74.

Haruya Suzuki, Yuto Miyauchi, Kazuki Akiyama, Tomoyuki Kajiwara, Takashi Ninomiya, Noriko Takemura, Yuta Nakashima, and Hajime Nagahara. 2022. A Japanese Dataset for Subjective and Objective Sentiment Polarity Classification in Micro Blog Domain. In *Proceedings of the 13th International Conference on Language Resources and Evaluation*, pages 7022–7028.

Rachael Tatman. 2017. Gender and Dialect Bias in YouTube's Automatic Captions. In *Proceedings of the First ACL Workshop on Ethics in Natural Language Processing*, pages 53–59.

Ryoko Tokuhisa, Kentaro Inui, and Yuji Matsumoto. 2008. Emotion Classification Using Massive Examples Extracted from the Web. In *Proceedings of the 22nd International Conference on Computational Linguistics*, pages 881–888.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is All you Need. In *Advances in Neural Information Processing Systems*, pages 5998–6008.

Maximilian Wich, Hala Al Kuwatly, and Georg Groh. 2020. Investigating Annotator Bias with a Graph-Based Approach. In *Proceedings of the Fourth Workshop on Online Abuse and Harms*, pages 191–199.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Remi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander Rush. 2020. Transformers: State-of-the-Art Natural Language Processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45.