Daisy at WASSA 2024 Empathy and Personality Shared Task: A Quick Exploration on Emotional Pattern of Empathy and Distress

Rendi Chevi
MBZUAI
rendi.chevi@mbzuai.ac.ae

Alham Fikri Aji
MBZUAI
alham.fikri@mbzuai.ac.ae

Abstract

When we encountered upsetting or tragic situations involving other people, we might feel certain emotions that are congruent, though not necessarily identical, to what that person might went through. These kind of *vicarious* emotions are what defined empathy and distress, they can be seen as a form of emotional response to other people in need. In this paper, we describe our participation in WASSA 2024 Shared Task 3 in predicting writer's level of empathy and distress from their personal essays. We approach this task by assuming one's level of empathy and distress can be revealed from the emotional patterns within their essay. By extracting the emotional patterns from essays via an emotion classifier, we regress the empathy and distress levels from these patterns. Through correlation and model explainability analysis, we found that there are similar set of emotions, such as *sadness* or *disappointment*, and distinct set of emotions, such as *anger* or *approval*, that might describe the writer's level of empathy and distress. We hope that our approach and findings could serve as a basis for future work that try to model and explain empathy and distress from emotional patterns.

1 Introduction

Some of us, in some situation, have the ability to infer other people's psychologically real, internal state (Zaki and Ochsner, 2012), such as their emotions or intentions. This ability can result in us experiencing "vicarious emotions", which are emotions that we feel when something happen to someone else (Wondra and Ellsworth, 2015).

Empathy and Distress as Emotions. Empathy and distress can be seen as a form of vicarious emotions, specifically elicited in response to perceiving other people in need (Batson et al., 1987). One of the main differentiator between the two is in the type of emotional response people tend to express when being empathic or distressed. Empathy tend to be more associated with feeling *compassionate*, *tender*, or *warm* (Batson et al., 1981, 1987), while distress tend to be more associated with feeling *alarmed*, *worried*, or *troubled* (Batson et al., 1987).

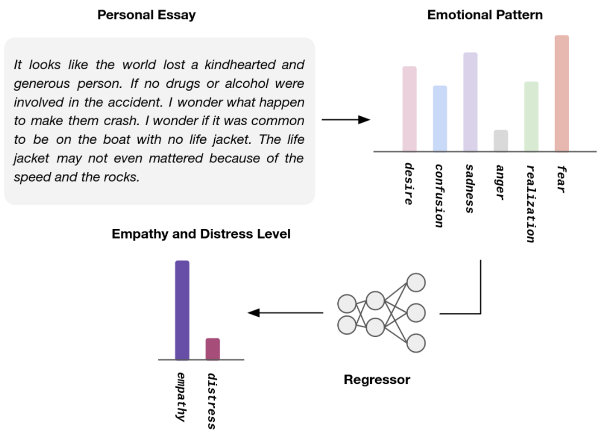

Figure 1: Overview of our approach. We extract emotional patterns from personal essay via an emotion classifier, and see whether we could regress and analyze the level of empathy and distress from these emotional patterns.

One could measure people's level of empathy and distress toward certain stimuli with emotional response-based scale following (Batson et al., 1987). This measure has also been used in (Buechel et al., 2018; Omitaomu et al., 2022), which is the basis of WASSA Shared Task dataset in the past years (Giorgi et al., 2024; Barriere et al., 2023).

Our Contributions. In our participation for WASSA 2024 *Shared Task 3: Empathy Prediction*, we intend to explore the task of predicting the writer's level of empathy and distress with almost exclusively leveraging the emotional pattern conveyed in their personal essay.

We are interested to see whether people's emotional response related to empathy and distress are reflected within their personal essay. If so, we

Proceedings of the 14th Workshop on Computational Approaches to Subjectivity, Sentiment, & Social Media Analysis, pages 420–424
August 15, 2024

might be able to analyze a set of similar or distinct emotional pattern attributed to empathy and distress. Then predict the latter from the former. Our main contributions in this paper are summarized as follows:

- We extracted the emotional pattern conveyed in personal essays via an emotion classifier. We found that there are similar and distinct set of emotions correlated with empathy and distress.
- We experimented to fit a regressor to predict the level of empathy and distress from emotional patterns. Though it performed decently on the evaluation set, it still can't generalize well to the test set.
- We analyzed the regressor using SHAP (SHapley Additive exPlanations) analysis (Lundberg and Lee, 2017). We found that there are similar and distinct set of emotions impacting the regressor's predictions.
- Lastly, we outline the limitations from this quick and early exploration, which we hope would serve as a basis for future work to model and explain empathy and distress from emotional patterns.

2 System Description

2.1 Preliminaries

Let's first define our dataset comprising tuples of $\{x_i, e_i, d_i\}_{i=1:n}$, where x_i is a personal essay, and e_i and d_i are the essay writer's level of empathy and distress (Batson et al., 1987), on a 7-point scale (1 = Not-at-all, 7 = Extremely).

As we would also like to incorporate emotional patterns conveyed in x_i, we define a set of emotion labels conveyed in x_i as $\mathbf{v}_i = \{v_i^j\}_{j=1:m}$. This can include a set of primary emotions (e.g. *sadness*, *joy*) (Ekman, 1992; Plutchik, 1984) or more complex ones (e.g. *bittersweetness*, *grief*) (Plutchik, 1984; Cowen et al., 2019; Demszky et al., 2020). Since this information is unavailable and out-of-scope of the WASSA 2024 dataset, we would have to extract it on our own.

2.2 Extracting Emotional Pattern from Personal Essay

To extract emotional patterns from each essay x_i, we utilize an open-source multi-label emotion classifier [1], based on RoBERTa-Base (Liu et al., 2019) finetuned on GoEmotion dataset (Demszky et al., 2020). Model trained on this dataset is suitable as a proxy for identifying the writer's emotions, as the annotators of GoEmotion were asked to identify emotions expressed by the writer. It also provides emotion labels based on semantic space taxonomy (Cowen et al., 2019) with 27 emotions and 1 neutral labels, which offers a more fine-grained and diverse emotion classification.

To account for varying tones and emotions within a personal essay, we segment the essay into several overlapping chunks, predict emotions for each segment with the emotion classifier, and then average the predictions to obtain the final emotional pattern for that particular essay.

2.3 Predicting Empathy and Distress from Emotional Pattern

After we obtain the emotional pattern $\mathbf{v}_i$ for each x_i, we fit a regression model to learn the mapping between the $\mathbf{v}_i$ and $\{e_i, d_i\}$. We are experimenting with 2 regressors: (1) Support Vector Regression (SVR), and (2) neural network-based regressor of 5 layers of perceptron with leaky ReLU activation in each subsequent layer. Both regressors takes in $\mathbf{v}_i$ and outputting $\{e_i, d_i\}$.

3 Results and Analysis

3.1 Correlation between Emotions, Empathy, and Distress

Before we get into predicting empathy and distress from emotional pattern, we are interested to see whether there is a notable linear correlation between the extracted emotional pattern, $\mathbf{v}_i$, and the level of empathy and distress, $\{e_i, d_i\}$.

Figure 2 shows the Pearson correlation the extracted emotional pattern from essays and the writer's level of empathy and distress. We can see that there are a set of similar and distinct emotions correlated with empathy and distress. For instance, both empathy and distress seem to be positively correlated with *caring*, *desire*, and *optimism*. But, we can also see that empathy is negatively correlated with *anger* while distress is the opposite. Distress seems to be negatively correlated with *grief*, while empathy is the opposite.

Though we can see some notable set of emotions correlated with empathy and distress, the correlations are independently weak. But, we can't fully expect there would be a single, independent emotion that linearly correlates with empathy and

[1] https://huggingface.co/SamLowe/roberta-base-go_emotions

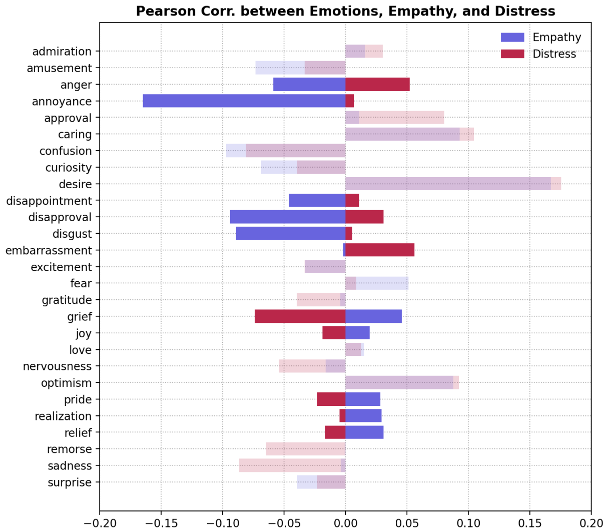

Figure 2: Pearson correlation between extracted emotional patterns from essays and empathy and distress. Emotions in opaque are ones that oppositely correlate with empathy and distress, while the transparent ones are correlated in the same direction.

distress. Even in (Batson et al., 1987), these vicarious emotions are associated and measured with multiple emotional adjectives. In the next section, we would report our result from fitting a regressor that learns a more complex, non-linear mapping between the extracted emotional patterns and empathy and distress.

3.2 Empathy and Distress Predictions from Emotional Patterns

Table 1 shows the results on empathy and distress prediction from emotional pattern, using the regressor described in Section 2.3. In predicting empathy from emotional pattern, we can see that the model performs decently across the unseen evaluation set, but had a notable performance drop in the test set which indicate overfitting. When predicting distress, though the regressor performs reasonably well in the evaluation set, it can't seem to generalize well into the test set.

Performance Drop in Test Set. It should be noted that we haven't done extensive efforts to improve the regressor's performance through regularization, hyperparameters tuning, or other methods, which might alleviate the overfitting in the test set.

We also hypothesized that maybe to approach this emotional pattern-based empathy-distress task, we should be taking other variables such as the stimuli and the writer's person factors (e.g. sociodemographics or trait empathy information) into account.

		Pearson Corr. ↑		
Model		Train Set	Eval Set	Test Set
SVR	Empathy	0.657	0.446	0.296
	Distress	0.625	0.416	0.165
	Avg.	0.641	0.431	0.231
MLP	Empathy	0.523	0.484	0.344
	Distress	0.488	0.515	0.082
	Avg.	0.506	0.499	0.213

Table 1: Empathy and distress prediction results from the regressors across dataset split.

As right now, we assume that all emotional patterns was elicited by the same stimuli and we don't consider how each person based on their person factors might have differences in the way they express empathy or distress from emotional sense.

3.3 Emotional Importance when Predicting Empathy and Distress

To further analyze the set of emotions that the regressor consider when predicting the value of empathy and distress, we conducted SHAP (Lundberg and Lee, 2017) analysis into the model. We conduct the analysis on the SVM regressor, as it is the one which has better trade-off in the empathy and distress prediction.

What Emotions are Impacting the Regressor's Prediction? Figure 3 and 4 shows the top-10 emotions that are impacting the model prediction the most. Each point represent an unseen sample from our evaluation and test dataset, its color tells the emotion level, and its position tells how much of that emotion pushes the model's prediction to the left (less empathic or distressed) or to the right (more emphatic or distressed).

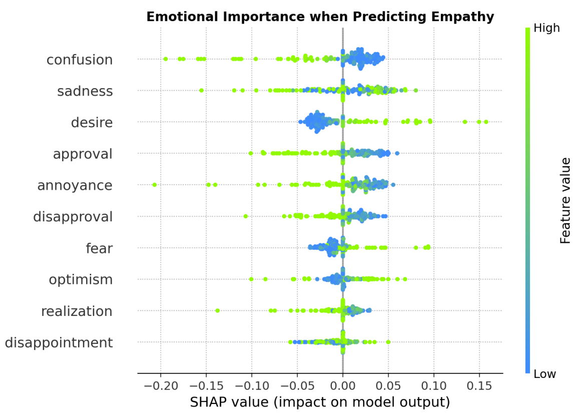

Figure 3: Emotional importance plot of the regressor predicting empathy.

In the empathy prediction, higher level of predicted empathy seems to be characterized by the lower the level of *confusion*, *approval*, *annoyance*, and *disapproval* and the higher level of *desire*, *fear*, and *optimism*. Interestingly, higher level of predicted distress could also be explained by similar level of emotions such as *desire*, *confusion*, and *fear*. But, the higher prediction of distress is also distinctly marked by *anger*, *curiosity*, *realization*, and surprisingly *caring*.

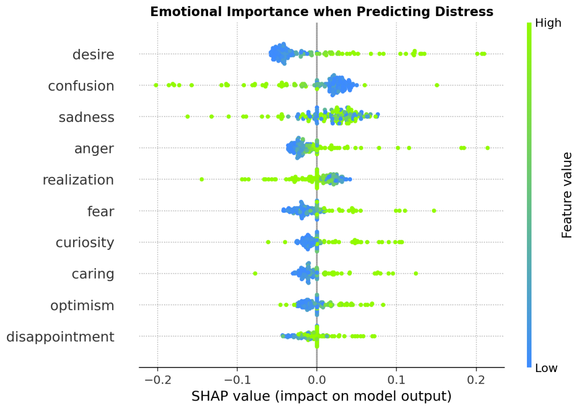

Figure 4: Emotional importance plot of the regressor predicting distress.

Another interesting observation is that, the level of *sadness* and *disappointment* doesn't seem to polarize the model prediction. Regardless of its level, the presence of *sadness* and *disappointment* tend to increase the predicted level of empathy and distress. Which makes sense, as the stimuli in the dataset are all involving upsetting and tragic news, where intuitively, people tend to vicariously share an expression of *sadness* and *disappointment* towards it.

4 Conclusion

In this paper, we describe our participation in the WASSA 2024 Shared Task 3 on predicting empathy and distress from personal essay. We approach this task by assuming that one's level of empathy and distress can be revealed from the emotional patterns expressed within their essay. Through extracting the emotional pattern in each personal essay, we directly fit a regressor into the emotional pattern and empathy-distress pairs. Our analysis shows that there are a set of similar and distinct emotions in the dataset and model's prediction that could describe empathy and distress. We hope that our approach and findings would serve as a basis for future work in modeling and explaining empathy and distress from emotional patterns.

Limitation and Future Work

This paper presents our quick and early exploration on the analysis and explanation of written expression of empathy and distress. That being said, we shouldn't yet be drawing any definitive conclusion at this stage, and there are notable limitations which are subjects for future works and ethical considerations. Those limitations are (not limited to): (1) **Emotion Classifier:** Modeling the perceiver-dependent or subjectivity of emotions are still an open problem, and using an emotion classifier that hasn't taken this into account may not reflect the writer's emotions in the essays and would lead to an inaccurate analysis and conclusion. Additionally, we should probably consider an emotion classifier (or regressor) that takes emotion ordinality and intensity into account, as the empathy and distress variables themselves are labeled ordinally. (2) **Regressor Performance:** Here, we haven't experimented to improve the performance of the regressors through hyperparameters tuning, regularization, or other methods, which may explain the problem of overfitting and generalizability in the test set. (3) **Explainability Analysis:** As we try to explain what emotions drive the model's prediction, our explanation would very much be model-dependent, improving the previous limitations is necessary to ensure we conduct a meaningful and definitive analysis.

Ethics Statement

It is important to acknowledge that not all people perceive and experience emotions the same, and not all people are able to adequately translate their emotions through essays or other kinds of modalities. We should not directly generalize any insights or findings derived from this study regarding people emotions, empathy, distress, and other psychological processes.

References

Valentin Barriere, João Sedoc, Shabnam Tafreshi, and Salvatore Giorgi. 2023. Findings of wassa 2023 shared task on empathy, emotion and personality detection in conversation and reactions to news articles. In *Proceedings of the 13th Workshop on Computational Approaches to Subjectivity, Sentiment, & Social Media Analysis*, pages 511–525.

C Daniel Batson, Bruce D Duncan, Paula Ackerman, Terese Buckley, and Kimberly Birch. 1981. Is empathic emotion a source of altruistic motivation? *Journal of personality and Social Psychology*, 40(2):290.

C Daniel Batson, Jim Fultz, and Patricia A Schoenrade. 1987. Distress and empathy: Two qualitatively distinct vicarious emotions with different motivational consequences. *Journal of personality*, 55(1):19–39.

Sven Buechel, Anneke Buffone, Barry Slaff, Lyle Ungar, and João Sedoc. 2018. Modeling empathy and distress in reaction to news stories. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 4758–4765, Brussels, Belgium. Association for Computational Linguistics.

Alan Cowen, Disa Sauter, Jessica L Tracy, and Dacher Keltner. 2019. Mapping the passions: Toward a high-dimensional taxonomy of emotional experience and expression. *Psychological Science in the Public Interest*, 20(1):69–90.

Dorottya Demszky, Dana Movshovitz-Attias, Jeongwoo Ko, Alan Cowen, Gaurav Nemade, and Sujith Ravi. 2020. Goemotions: A dataset of fine-grained emotions. *arXiv preprint arXiv:2005.00547*.

Paul Ekman. 1992. An argument for basic emotions. *Cognition & emotion*, 6(3-4):169–200.

Salvatore Giorgi, João Sedoc, Valentin Barriere, and Shabnam Tafreshi. 2024. Findings of wassa 2024 shared task on empathy and personality detection in interactions. In *Proceedings of the 14th Workshop on Computational Approaches to Subjectivity, Sentiment, & Social Media Analysis*.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Scott M Lundberg and Su-In Lee. 2017. A unified approach to interpreting model predictions. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 4765–4774. Curran Associates, Inc.

Damilola Omitaomu, Shabnam Tafreshi, Tingting Liu, Sven Buechel, Chris Callison-Burch, Johannes Eichstaedt, Lyle Ungar, and João Sedoc. 2022. Empathic conversations: A multi-level dataset of contextualized conversations. *Preprint*, arXiv:2205.12698.

Robert Plutchik. 1984. Emotions: A general psychoevolutionary theory. *Approaches to emotion*, 1984(197-219):2–4.

Joshua D Wondra and Phoebe C Ellsworth. 2015. An appraisal theory of empathy and other vicarious emotional experiences. *Psychological review*, 122(3):411.

Jamil Zaki and Kevin N Ochsner. 2012. The neuroscience of empathy: progress, pitfalls and promise. *Nature neuroscience*, 15(5):675–680.