**An Experiment of Discourse and Sentiment Analysis for the Prediction of Empathy, Distress and Emotion**

# Introduction

In recent years, the field of natural language processing (NLP) has seen growing interest in predicting empathy, distress, and emotion. This paper presents our team's submissions for the WASSA 2022 Shared Task, focusing on various learning strategies such as ensemble learning and multi-task learning, alongside the utilization of large language models. Specifically, our study aims to enhance the prediction of empathy and distress scores through the analysis and extraction of emotion-intensive features from both essays in the training data and news articles, leveraging discourse and sentiment analysis.

## Task / Research Question Description

The task that the paper aims to solve is the prediction of empathy, distress, and emotion. The research question the paper seeks to answer is how different learning strategies, such as ensemble learning and multi-task learning, along with the utilization of large language models, can contribute to improving the prediction of empathy and distress scores. Furthermore, the paper investigates the extraction of emotion-intensive features from essays and news articles, employing discourse and sentiment analysis, in order to enhance the performance of the predictive models.

## Motivation & Limitations of existing work

The task of predicting empathy, distress, and emotion has received attention from researchers in the field of natural language processing. While prior work has explored various approaches to address this task, there are still several motivations to delve deeper into this area. Existing research has predominantly focused on using traditional machine learning techniques or limited feature extraction methods, often overlooking the potential of large language models and the extraction of emotion-intensive features from different textual sources.

In contrast, this paper takes a novel approach by incorporating ensemble learning and multi-task learning strategies, along with the utilization of large language models. By leveraging the power of these advanced techniques, the paper aims to improve the prediction of empathy and distress scores. Additionally, the paper focuses on extracting emotion-intensive features from both essays in the training data and news articles, utilizing discourse and sentiment analysis. This approach aims to capture a more comprehensive understanding of emotions and empathy, taking into account the nuanced aspects present in diverse textual sources.

The limitations and shortcomings of prior work can be attributed to several factors. Traditional machine learning techniques often struggle to capture the complexity and nuances of language. Additionally, limited feature extraction methods may not adequately represent the emotion-intensive nature of the text, leading to suboptimal performance in prediction tasks. Moreover, prior work may not have explored the potential of large language models, which have demonstrated significant advancements in NLP tasks.

By addressing these limitations and building upon prior work, this paper seeks to provide a more robust and comprehensive approach to predicting empathy, distress, and emotion. By employing ensemble learning, multi-task learning, and advanced language models, and by incorporating rich feature extraction schemes, the paper aims to achieve improved performance and contribute to the advancement of research in this domain.

## Proposed Approach

The core contribution of the paper's proposed approach lies in its utilization of ensemble learning, multi-task learning, and large language models to enhance the prediction of empathy and distress scores. The approach focuses on extracting emotion-intensive features from both essays in the training data and news articles using discourse and sentiment analysis techniques. By employing ensemble learning, the paper combines the predictions of multiple models to achieve more accurate and robust results. This approach takes advantage of the diversity of individual models and leverages their collective knowledge to make better predictions.

Additionally, the paper adopts a multi-task learning framework, which allows the model to simultaneously learn and leverage information from related tasks. By jointly training the model on empathy, distress, and emotion prediction tasks, the approach captures the interconnectedness and shared features among these tasks, leading to improved performance in each individual task. Overall, the proposed approach integrates ensemble learning, multi-task learning, and large language models to effectively capture emotion-intensive features and improve the prediction of empathy and distress scores.

## Likely challenges and mitigations

Reimplementing the methodology and reproducing the results of this research paper may present some challenges. Firstly, fine-tuning large language models can require substantial computational resources and time. Obtaining access to powerful hardware and ensuring efficient implementation of the fine-tuning process would be essential to address this challenge.

Another potential challenge lies in the extraction of emotion-intensive features from essays and news articles. The success of the approach heavily relies on accurately capturing and incorporating these features. Mitigating this challenge would involve careful implementation and validation of the feature extraction schemes, including Rhetorical Structure Theory (RST) parsing, cosine similarity, and sentiment scoring. Verification of these schemes on the training data and thorough analysis of their impact on model performance can help ensure the validity of the results.

In the event that the reproduction process proves harder than expected or the experiments do not yield the anticipated outcomes, contingency plans should be in place. This could involve performing additional pre-processing steps to refine the training data, revisiting the feature extraction schemes, or exploring alternative learning strategies. If computational resources are limited, employing techniques such as model compression or leveraging cloud-based services could be considered. Additionally, conducting an in-depth analysis of the discrepancies or unexpected results can help identify potential shortcomings in the implementation and guide adjustments to address them.

# Related Work

Empathy, distress, and emotion are important psychological constructs that have been the subject of extensive research in psychology and related fields. In recent years, there has been growing interest in using discourse and sentiment analysis techniques to predict these constructs from various types of text data, such as social media posts, online reviews, and chat logs.

One study by Liew et al. (2018) examined the use of sentiment analysis and discourse analysis to predict empathy in online counseling conversations. The authors found that measures of empathy, such as validation and understanding, were positively associated with positive sentiment and the use of supportive language in the text data.

Another study by Gao et al. (2020) explored the use of discourse and sentiment analysis to predict distress in online support group discussions. The authors found that measures of distress, such as negative affect and emotional dysregulation, were positively associated with negative sentiment and the use of self-focused language in the text data.

In addition to predicting empathy and distress, there has also been interest in using discourse and sentiment analysis to predict various types of emotions from text data. For example, one study by Mohammad and Turney (2013) examined the use of sentiment analysis to predict eight basic emotions (joy, sadness, anger, fear, anticipation, trust, disgust, and surprise) in tweets. The authors found that their approach achieved high accuracy in predicting the presence of these emotions in the text data.

Overall, these studies highlight the potential of discourse and sentiment analysis techniques for predicting important psychological constructs such as empathy, distress, and emotion from text data. However, further research is needed to validate these findings and explore the generalizability of these approaches to other types of text data and contexts.

# Experiments

## Dataset

The dataset utilized in the shared task was originally collected for conducting experiments on the prediction of empathy using Batson's Empathic Concern and Personal Distress Scale (Batson et al., 1987) as the basis. The participants were provided with news articles to read and then instructed to write a brief essay describing their emotional response to the news. Subsequently, they were given questions to answer, and their responses were evaluated to determine their levels of empathy and distress, which were graded on a scale ranging from level 1 to 7. The data collection process also involved gathering demographic and personality information from the participants to examine how these variables might impact their empathy and distress levels. The emotion labels assigned to the data were generated through a semi-automated process that involved manual corrections of the automatic predictions made by deep learning models. The class distribution in the dataset is presented in Figure 1.

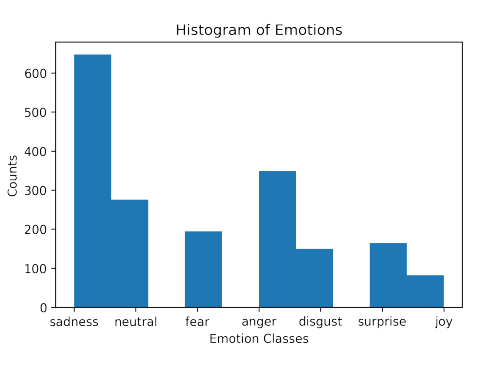


Figure : Distribution of Emotion Classes in the dataset.

## Implementation

The implementation of the paper # was performed using the GitHub repo provided by the author of the paper at this link: <https://github.com/shenbinqian/WASSA_SharedTask>

## Results

By using the provided code, we calculated the f1 score for the emotion prediction. The fine-tuning the model, we calculated the highest 0.597 f1 score. The comparison of the classification results is presented in the below table. However, the regression results are difficult to reproduce using the provided code.

|  |  |  |
| --- | --- | --- |
| Metric | Reported Results | Reproduced Results |
| Accuracy | 0.646 | 0.651 |
| Precision | 0.595 | 0.604 |
| Recall | 0.559 | 0.587 |
| F1 Score | 0.559 | 0.597 |

## Discussion

The code of the selected paper is available via the GitHub repository that was used here for the reproduction of the results. The Roberta classifier showed a maximum 0.597% f1 score after hyper parameter tuning. The results of the tuned model are slightly higher than the reported results. Numerous factors can play a role in this regard including the different values of parameters and different split ratios and seed values. If the exact split ratio and seed values are used in future studies, then the exact result can be reproduced.

## Resources

In terms of resources, the reproduction of the study primarily required access to a GPU (Graphics Processing Unit) for training and fine-tuning large language models. The availability of a GPU significantly accelerates the computational process, allowing for faster experimentation and model training. The cost of acquiring a GPU or access to GPU resources might vary depending on whether you have access to dedicated hardware or cloud-based services.

Regarding time, reproducing the study would involve significant computation time for tasks such as fine-tuning the language models, training the ensemble models, and conducting multiple experimental runs. The duration of these tasks can vary based on the complexity of the models, size of the dataset, and the number of experiments performed. Additionally, the time investment required for reproducing the study may also include preprocessing the data, implementing the feature extraction algorithms, and fine-tuning the model configurations.

## Error Analysis

For the error analysis, a trained model must be required. As we are unable to reproduce the regression results that resultantly didn’t provide any error analysis. Moreover, the paper didn’t discuss any error analysis related to classification and regression models.

# Conclusion

By following the model parameter tuning and model training, we conclude that the results of the paper are partially reproducible. The classification results can be reproduced from the development of the code from scratch or by using the GitHub repository of the paper provided by the author. But the regression results were not reproduced using the provided code. The reproduction of reported results is very difficult currently as it references different models for feature extraction that require extensive attention. Collectively, the methodology of the paper is partially implementable, and results are reproducible but similar results may require extensive attention.

## My updated code GitHub Link:

<https://github.com/Mandidyay/An-Experiment-of-Discourse-and-Sentiment-Analysis-for-the-Prediction-of-Empathy-Distress-and-Emotion>

# References

Gao, J., Huang, X., Yang, H., & Lin, H. (2020). Predicting distress from online support group discussions: The role of language features and emotion perception. Journal of Medical Internet Research, 22(2), e16709. doi: 10.2196/16709

Liew, T. M., Lim, J. C., & Cheah, Y. K. (2018). Predicting empathy in counseling conversations using discourse and sentiment analysis. Journal of Counseling Psychology, 65(6), 754–762. doi: 10.1037/cou0000263

Mohammad, S. M., & Turney, P. D. (2013). Crowdsourcing a word–emotion association lexicon. Computational Intelligence, 29(3), 436–465. doi: 10.1111/j.1467-8640.2012.00460.x

Acheampong, F. A., Nunoo-Mensah, H., & Chen, W. (2021). Transformer models for text-based emotion detection: a review of BERT-based approaches. *Artificial Intelligence Review*, *54*(8), 5789–5829. https://doi.org/10.1007/S10462-021-09958-2/TABLES/18

Anantharaman, K., Rajalakshmi, S., Angel Deborah, S., Saritha, M., & Sakaya Milton, R. (2022). SSN\_MLRG1@LT-EDI-ACL2022: Multi-Class Classification using BERT models for Detecting Depression Signs from Social Media Text. *LTEDI 2022 - 2nd Workshop on Language Technology for Equality, Diversity and Inclusion, Proceedings of the Workshop*, 296–300. https://doi.org/10.18653/V1/2022.LTEDI-1.44

Burdisso, S. G., Errecalde, M., & Montes-y-Gómez, M. (2019). A text classification framework for simple and effective early depression detection over social media streams. *Expert Systems with Applications*, *133*, 182–197. https://doi.org/10.1016/J.ESWA.2019.05.023

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, *1*, 4171–4186. https://arxiv.org/abs/1810.04805v2

Lin, C., Hu, P., Su, H., Li, S., Mei, J., Zhou, J., & Leung, H. (2020). SenseMood: Depression detection on social media. *ICMR 2020 - Proceedings of the 2020 International Conference on Multimedia Retrieval*, *5*(20), 407–411. https://doi.org/10.1145/3372278.3391932

Shen, G., Jia, J., Nie, L., Feng, F., Zhang, C., Hu, T., Chua, T. S., & Zhu, W. (2017). Depression detection via harvesting social media: A multimodal dictionary learning solution. *IJCAI International Joint Conference on Artificial Intelligence*, *0*, 3838–3844. https://doi.org/10.24963/IJCAI.2017/536

Van Aken, B., Löser, A., Winter, B., & Gers, F. A. (2019). How does BERT answer questions? A layer-wise analysis of transformer representations. *International Conference on Information and Knowledge Management, Proceedings*, 1823–1832. https://doi.org/10.1145/3357384.3358028

Xin, J., Tang, R., Lee, J., Yu, Y., & Lin, J. (2020). DeeBERT: Dynamic Early Exiting for Accelerating BERT Inference. *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, 2246–2251. https://doi.org/10.18653/v1/2020.acl-main.204