Leveraging Emotion-specific Features to Improve Transformer Performance for Emotion Classification

Atharva Kshirsagar*[1], **Shaily Desai***[1], **Aditi Sidnerlikar**[1], **Nikhil Khodake**[1] and **Manisha Marathe**[2]

[1,2]Department of Computer Engineering, PVG's COET, Affiliated to Savitribai Phule Pune University, India.

[1]{atharvakshirsagar145, shaily.desai21,
[1]sidnerlikaraditi6, nikhilkhodake2002} @gmail.com
[2]mvm_comp@pvgcoet.ac.in

Abstract

This paper describes team PVG's AI Club's approach to the Emotion Classification shared task held at WASSA 2022. This Track 2 subtask focuses on building models which can predict a multi-class emotion label based on essays from news articles where a person, group or another entity is affected. Baseline transformer models have been demonstrating good results on sequence classification tasks, and we aim to improve this performance with the help of ensembling techniques, and by leveraging two variations of emotion-specific representations. We observe better results than our baseline models and achieve an accuracy of 0.619 and a macro F1 score of 0.520 on the emotion classification task.

1 Introduction

Rapid growth in the availability of human-annotated text documents has led to an increase in methodologies for tasks such as classification, clustering and knowledge extraction. A multitude of sources have enabled public access to structured and semi-structured data comprising of news stories, written repositories, blog content, among countless other roots of information. (Bostan and Klinger, 2018) showed that the task of emotion classification has emerged from being purely research oriented to being of vital importance in fields like dialog systems, intelligent agents, and analysis and diagnosis of mental disorders.

Humans themselves sometimes find it tough to comprehend the various layers of subtlety in emotions, and hence there has been only a limited amount of prior research revolving around emotion classification. It has been noted that larger deep learning models can also find it quite challenging to fully grasp the nuances and underlying context of human emotion.

With the advent of Transformer (Vaswani et al., 2017) models, there has been an increase in performance for emotion classification of text-based models. Most transformer-based language models (Devlin et al., 2018; Raffel et al., 2019; Radford et al., 2018) are pretrained on various self-supervised objectives. Combining transformer based sentence representations with domain-specialised representations for improving performance on the specific task has been successfully used in across many NLP domains (Peinelt et al., 2020; Poerner et al., 2020; Zhang et al., 2021). Building on these foundations, we propose a similar approach to the task of Emotion classification.

In this paper, we posit a solution to the WASSA 2022 Shared Task on Empathy Detection, Emotion Classification and Personality Detection, specifically Track-2, emotion classification. We propose a hybrid model where we combine information from various entities to create a rich final representation of each datapoint, and the observed results show promise in combining the Transformer output with the emotion-specific embeddings and NRC features.

The rest of the paper is organized in the following manner: Section 2 offers an overview into the dataset on Empathetic concern in news stories, Section 3 goes in depth about our proposed methodology with subsections describing the individual constituent modules. Section 4 explains the experimental and training setup along with the baselines used; Section 5 elucidates the observed results, and Section 6 concludes this study.

2 Dataset

The dataset provided by the organizers consists of 1860 essays in the training set, 270 in the dev set and 525 in the test set. Each of these essays has been annotated for empathy and empathy scores, distress and distress scores, emotion, personality feature and interpersonal reactivity features. Since

*Equal Contribution

Proceedings of the 12th Workshop on Computational Approaches to Subjectivity, Sentiment & Social Media Analysis, pages 245 - 249

this paper describes an approach only to the Emotion classification task, we shall only describe the data for said subtask. Each essay has been assigned an emotion class similar to classes in (Ekman, 1992). Table 1 provides a description on the training, validation and testing subset, and Figure 1 shows the distribution of the training data among the various emotion classes.

Set	Essays
Training	1860
Validation	270
Testing	525
Total	2655

Table 1: Total datapoints for every set

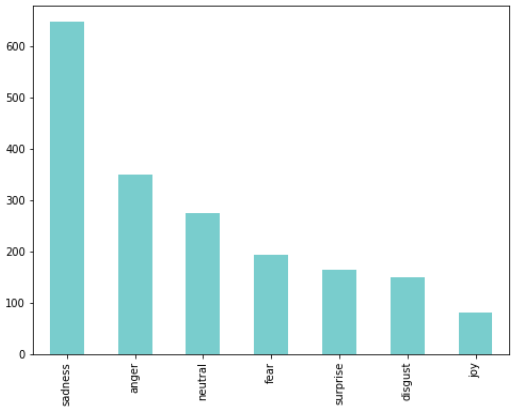

Figure 1: Distribution of the various classes among the Training Dataset

3 Methodology

3.1 RoBERTa

We make use of the pretrained RoBERTa base model (Liu et al., 2019) for this task. RoBERTa provides contextualized essay-level representations which can capture context sensitive information better than static representations. For each essay E in our corpus, we obtain a 768 dimensional representation R, encoded using the CLS token in the final hidden layer of the RoBERTa base model. We further process this representation R with Linear and Dropout layers before concatenating it with our emotion-specific representations.

$$R = RoBERTa(E) \in \mathbb{R}^{d_1} \quad (1)$$

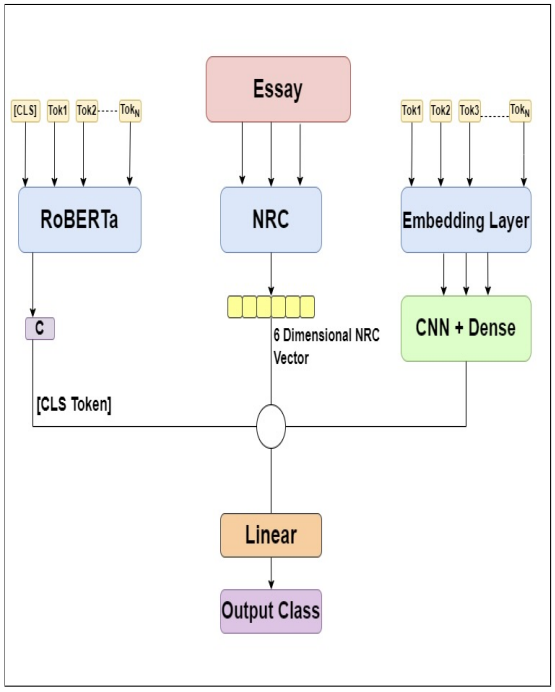

Figure 2: Model Architecture

3.2 Emotion-Enriched Word Embeddings(EWE)

(Labutov and Lipson, 2013; Bansal et al., 2014),argue that the effectiveness of word embeddings is highly task dependent. To obtain word embeddings specific for emotion classification, we used the emotion-enriched embeddings from (Agrawal et al., 2018). The weight matrix was made by mapping the vocabulary from our dataset to the 300 dimensional corresponding vector in the pre-trained embedding file. Each essay was mapped to the embedding matrix into a final representation shape of (100,300). This representation was passed through 2 Conv1d and 2 Maxpool layers to obtain a 16 dimensional feature vector $C \in \mathbb{R}^{d_2}$.

3.3 NRC Representation

The NRC emotion intensity lexicon (Mohammad, 2018) is a collection of close to $10,000$ words associated with a distinct real valued intensity score assigned for eight basic emotions. Incorporating this lexicon in classification tasks has been proven to boost performance (Kulkarni et al., 2021). Of the 8 basic emotions in the lexicon, 6 emotions-anger, joy, sadness, disgust, fear and surprise coincide with the given dataset and hence lexical features for only these features were considered. For every essay in the dataset, we calculate the value for one emotion by summing the individual scores for

Model	Accuracy		Macro-F1 Score	
	Training	Validation	Training	Validation
Vanilla RoBERTa	0.601	0.540	0.513	0.452
RoBERTa + EWE	0.684	0.608	0.561	0.499
RoBERTA + NRC + EWE	**0.693**	**0.619**	**0.618**	**0.520**

Table 2: Resulting metrics on baseline models as compared to our methodology

every word W in the essay that occurs in the NRC lexicon. We then create a six dimensional vector N corresponding to that essay which consists of the scores of the emotions in our dataset.

For a datapoint E, the six values of $S_{emotion}$ and the feature vector N was constructed in the following manner:

$$S_{emotion} = \sum W_{emotion}(W \in E) \quad (2)$$

$$N = [S_{anger}; S_{joy}; .....; S_{surprise}] \in \mathbb{R}^{d_3} \quad (3)$$

3.4 Combined Representation and Classification

The feature vectors obtained from the RoBERTa (R), Emotion-Enriched Embeddings (C) and NRC (N)were concatenated to obtain the final representation (F).

$$F = [R; C; N] \in \mathbb{R}^{d_1+d_2+d_3} \quad (4)$$

This representation is then passed through a single Linear layer with the Softmax activation. Figure 2 depicts the model architecture in detail.

4 Experimental Setup

4.1 Data Preparation

Standard text cleaning steps like removing numbers, special characters, punctuation, accidental spaces, etc. were applied to each essay in the corpus. Stopwords were removed using the `nltk` (Loper and Bird, 2002) library. Every essay was tokenized to a maximum length of 100, and essays larger than this length were truncated. No standardization was done in the case of NRC scores, as we wanted to feed our model a vector of raw emotion-intensity scores for each of the six emotions considered in our NRC representation.

4.2 Training Setup

We used the pretrained 'roberta-base' model from the Huggingface `Transformers` [1] library. All other modules used in our methodology were built using PyTorch. As observed by (Kulkarni et al., 2021), we also found that the Hyperbolic Tangent(Tanh) activation function worked better than ReLU, and hence we used the Tanh activation for all layers in our model. The model was trained using an AdamW optimizer (Loshchilov and Hutter, 2019) with a learning rate of 0.001 and beta values set to $\beta_1 = 0.9$, $\beta_2 = 0.99$ and the loss used was cross entropy loss. Additionally, early stopping was used if the validation loss does not decrease after 10 successive epochs. The batch size was set to 64 for both Baseline models as well as the proposed model. A single Nvidia P100-16GB GPU provided by Google Colab was used to train all models.

4.3 Baselines

Our goal in this work is to examine if concatenating emotional-specific features to pre-existing transformer models leads to an increase in the emotion classification performance of these models. Hence, we compare our proposed methodology to the vanilla RoBERTa model, as well as RoBERTa + Ewe for the emotion classification subtask.

5 Results and Discussion

The results for the emotion prediction task on the validation set are given in Table 2. There was no use of validation data during the training process, and the provided validation data was used as unseen testing data to benchmark the models.The official metric for Track 2 of the shared task was the macro F1 score. To ensure fair comparison, the validation set results have been averaged over 3 runs for each model. The proposed model shows a 7% increase in macro F1 scores and 8% increase in ac-

[1] https://huggingface.co/transformers/

curacy over the vanilla RoBERTa model. The proposed model also shows the effectiveness of adding the NRC representations described in section 3.3 as it performs slightly better than the RoBERTa + Emotion Enriched word embeddings model. We attribute this increase in performance to the task-specific representations of essays used in our system. During the training process, it was observed that the performance of all models was highly susceptible to how they were initialized, and we received a large range of results across different seeds. As a result, a true assessment of our method can only be made in comparison to baseline models with the same seed, as we have done in this study.

6 Conclusion

The goal of this study was to examine and enhance the performance of transformer models using only the Empathetic Concern in News Stories dataset that was provided to us, with the prospective of testing our method on a bigger dataset in the future. We proposed a model ensemble which combined the transformer feature vector with the emotion-intensive word embeddings along with the word-specific features obtained from the NRC lexicon. We demonstrate results that outperform the baseline vanilla RoBERTa model, and attest that combining domain-specific features can indeed improve performance on a task as involute as emotion classification.

References

Ameeta Agrawal, Aijun An, and Manos Papagelis. 2018. Learning emotion-enriched word representations. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 950–961, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Mohit Bansal, Kevin Gimpel, and Karen Livescu. 2014. Tailoring continuous word representations for dependency parsing. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 809–815, Baltimore, Maryland. Association for Computational Linguistics.

Laura-Ana-Maria Bostan and Roman Klinger. 2018. An analysis of annotated corpora for emotion classification in text. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 2104–2119, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

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

Atharva Kulkarni, Sunanda Somwase, Shivam Rajput, and Manisha Marathe. 2021. PVG at WASSA 2021: A multi-input, multi-task, transformer-based architecture for empathy and distress prediction. In *Proceedings of the Eleventh Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 105–111, Online. Association for Computational Linguistics.

Igor Labutov and Hod Lipson. 2013. Re-embedding words. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 489–493, Sofia, Bulgaria. Association for Computational Linguistics.

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*, abs/1907.11692.

Edward Loper and Steven Bird. 2002. Nltk: The natural language toolkit. In *Proceedings of the ACL-02 Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics - Volume 1*, ETMTNLP '02, page 63–70, USA. Association for Computational Linguistics.

Ilya Loshchilov and Frank Hutter. 2019. Decoupled weight decay regularization. In *International Conference on Learning Representations*.

Saif Mohammad. 2018. Word affect intensities. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).

Nicole Peinelt, Dong Nguyen, and Maria Liakata. 2020. tBERT: Topic models and BERT joining forces for semantic similarity detection. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7047–7055, Online. Association for Computational Linguistics.

Nina Poerner, Ulli Waltinger, and Hinrich Schütze. 2020. E-BERT: Efficient-yet-effective entity embeddings for BERT. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 803–818, Online. Association for Computational Linguistics.

Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. 2018. Improving language understanding by generative pre-training.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. 2019. Exploring the limits of transfer learning with a unified text-to-text transformer. *arXiv preprint arXiv:1910.10683*.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.

Jinpeng Zhang, Baijun Ji, Nini Xiao, Xiangyu Duan, Min Zhang, Yangbin Shi, and Weihua Luo. 2021. Combining static word embeddings and contextual representations for bilingual lexicon induction. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 2943–2955, Online. Association for Computational Linguistics.