ACL 2024 Bangkok, Thailand

WASSA 2024 Shared Task

Paper ID: WASSA-EMP-10

hyy33 at WASSA 2024 Empathy and Personality Shared Task: Using the CombinedLoss and FGM for Enhancing BERT-based Models in Emotion and Empathy Prediction from Conversation Turn

Huiyu Yang, Liting Huang, Tian Li, Nicolay Rusnachenko, Huizhi Liang*

Newcastle University, Newcastle Upon Tyne, England

Introduction

Emotion detection and empathy analysis are important and inevitable topics with great application potentials. To provide more insights, **WASSA 2024 Shared Task** focuses on Empathy Detection and Emotion Classification and Personality Detection.

We propose a solution towards Track 2: Empathy and Emotion Prediction in Conversations Turns (CONV-turn), predicting the **Emotion, Emotion Polarity and Empathy** according to turn-level information during conversations

– To achieve this goal:

We adopt **BERT** and its variation of **DeBERTa** as base models, and **fine-tuned** them on task-oriented data with **adversarial training by Fast Gradient Method (FGM).** We also designed the **CombinedLoss**, which consisted of a structured contrastive loss and a Pearson loss.

– After submitting to the competition:

The **Segmented Mix-up** was proposed for data augmentation, and **boosting** was adopted as **ensemble** strategy. **Regression** experiments are further conducted.

The Dataset

In this task, participants are given **text information from conversations** between two users that read the same essay, which contains reaction to news articles where there is harm to a person or group.

– The dataset of Track 2 includes:

Training set: 11,166 samples

Development set: 990 samples

Test set: 2,061 valid samples

– Each sample consists:

Turn-level text information from single dialogue turns

The labels of perceived **Emotion, Emotional Polarity and Empathy**

Meta information of the speakers and the conversation

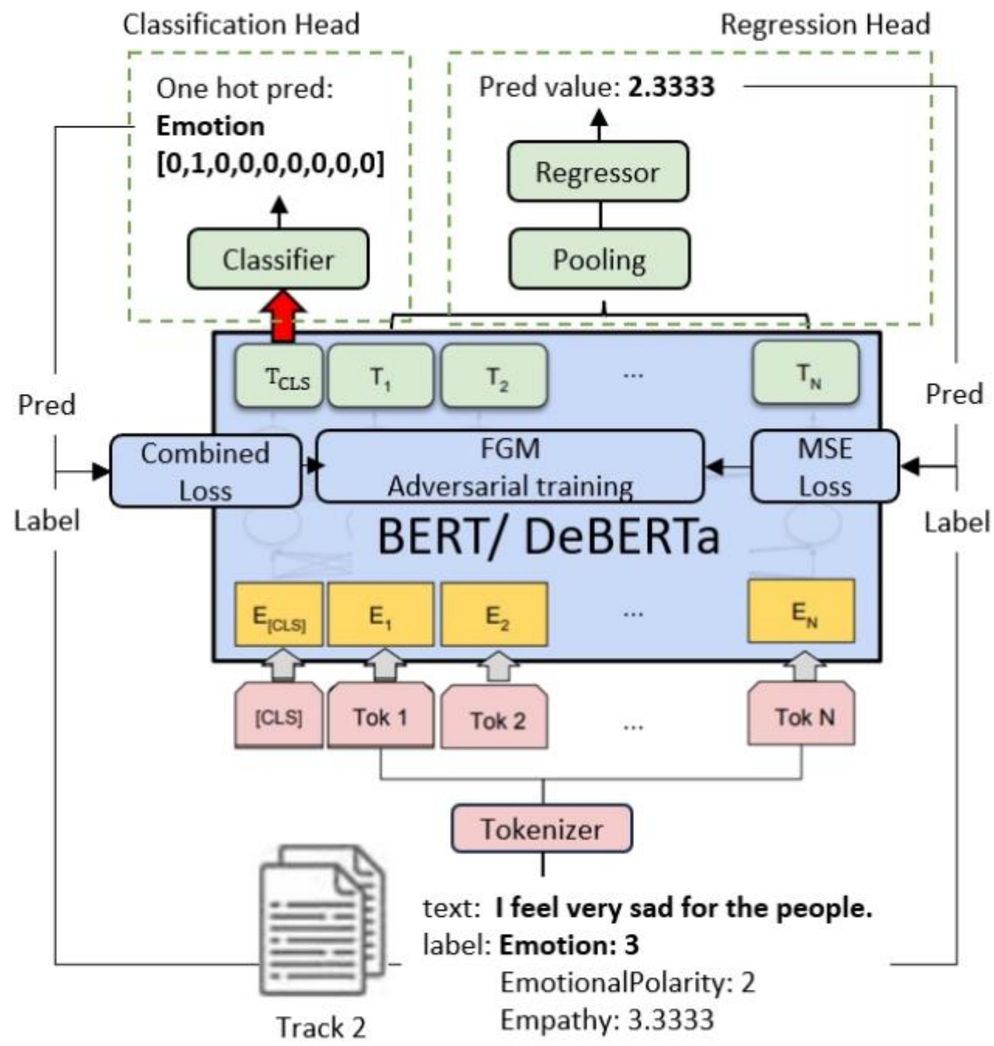
A Training Sample from Track 2
Text: I can't imagine just living in an area that is constantly being ravaged by hurricanes or earthquakes. I take my location for granted.
Label: Emotion: 3 EmotionalPolarity: 2 Empathy: 4.6667 SelfDisclosure: 3.3333
Other meta information: id: 3, article_id: 35, conversation_id: 1, turn_id: 3, speaker: "Person 2", person_id_1: "p019", person_id_2: "p012"

– The Evaluation Metric

Pearson correlation of the prediction sequence $\hat{y}$ and the ground truth sequence y

$$Corr_P(\hat{y}, y) = \frac{\sum_{i=1}^{n}\left(\frac{(\hat{y}_i - \bar{\hat{y}})}{\sigma_{\hat{y}}}\frac{(y_i - \bar{y})}{\sigma_y}\right)}{n}$$

The Methodology



– Fine-tuned BERT and DeBERTa

BERT: bert-base-uncased, 110M parameters with 12 encoder layers

DeBERTa: deberta-base, 390M parameters.

– The CombinedLoss

Different from commonly-used loss functions, we proposed the **CombinedLoss**

The Pearson correlation coefficient is used as a **regularization term**

$$L_{\text{total}} = L_{\text{loss}} + \lambda(1 - Corr_{Pear}(\hat{\mathbf{y}}, \mathbf{y}))$$

- L_{loss}: the structured contrastive loss for classification
- λ: the regularization coefficient
- $Corr_{Pear}(\hat{y},\ y)$: Pearson correlation coefficient between prediction and ground truth

– Adversarial Training with FGM

To improve its robustness and generalization, **FGM** is used as **adversarial training**

By maximizing $L(f_\theta(x+\delta))$, the most disturbing perturbation are introduced

The model is then trained to minimize the error, which helps it to be more robust

$$Obj = \min_{\theta} E(x, y)\left[\max L(f_\theta(x+\delta), y)\right]$$

- x: the input sample
- δ: the added perturbation for adversarial training
- f_θ: neural network function with θ as parameters

FGM computes the **most disturbing perturbation** through scaling the gradient

– Augmentation with the Segmented Mix-up

Mix-up is often used as **data augmentation**.

Mix-up without constraint can't generate meaningful samples => **Segmented Mix-up**

Samples are divided regarding their labels: the lower segment and the upper segment

Sample (x_i, y_i) is paired with (x_j, y_j) from the **same label segment**, they generate:

$$\tilde{x}_i = \mu x_i + (1-\mu)x_j$$

$$\tilde{y}_i = \mu y_i + (1-\mu)y_j$$

– Ensemble with Boosting

To build more accurate and robust system, **ensemble** is adopted with **boosting**

Weights are assigned regarding the accuracy of each model on development set

The model with the **most reliable prediction** has the **greatest impact**

Experiments and results

– Fine-tuned BERT and DeBERTa

The average results of fine-tuned **DeBERTa** is better than fine-tuned BERT

By adding the **CombinedLoss**, both models demonstrate performance gain

Adding **adversarial training using FGM** brings better overall performance

Model	Loss	FGM	Emo	EmoP	Emp	Avg
BERT	Cross-entropy	No	0.5867	0.6824	0.5703	0.6131
BERT	CombinedLoss	No	0.5921	0.6836	0.5803	0.6187
BERT	CombinedLoss	Yes	**0.6142**	**0.6899**	**0.5852**	**0.6298**
DeBERTa	Cross-entropy	No	0.6255	0.7281	0.5918	0.6485
DeBERTa	CombinedLoss	No	0.6348	0.7364	0.6042	0.6585
DeBERTa	CombinedLoss	Yes	**0.6399**	**0.7366**	**0.6064**	**0.6610**

– Ensemble and Augmentation

The **combined boosting** yields the best avg. result among non-augment models

Ensembling fine-tuned DeBERTas not always achieves the highest score

Augmentation with our **Segmented Mix-up** brings further improvement

Model	Ensemble	Augment	Emo	EmoP	Emp	Avg
BERT	Boosting	No	0.6521	0.7045	0.6069	0.6545
DeBERTa	Boosting	No	0.6470	0.7215	0.6112	0.6599
BERT, DeBERTa	Boosting	No	0.6485	0.7253	0.6140	0.6626
BERT, DeBERTa	Boosting	Mix-up	**0.6521**	**0.7334**	**0.6326**	**0.6727**

– Classification and Regression

The results of the fine-tuned DeBERTa in **different downsteam tasks**

The fined-tuned DeBERTa achieved slightly better performance in **regression task**

Model	Task	Emo	EmoP	Emp	Avg
DeBERTa	Classification	0.6399	0.7366	0.6064	0.6610
DeBERTa	Regression	**0.6409**	**0.7376**	**0.6105**	**0.6630**

Conclusion

– This paper presents our solution to **WASSA 2024 Track 2**, predicting **Emotion, Emotional Polarity and Empathy** using turn-level information.

– **BERT and DeBERTa** is fine-tuned with **adversarial training using FGM.** Models are trained with the **CombinedLoss**.

– The proposed method achieved **Pearson correlation of 0.581 for Emotion, 0.644 for Emotional Polarity and 0.544 for Empathy** on the test set, with the average value of 0.590 (**ranked 4th** among all teams).

– After the submission, **ensemble** with **boosting** method and **data augmentation** with **Segmented Mix-up** are adopted, which yield **even better results**: 0.6521 for Emotion, 0.7376 for Emotional Polarity, 0.6326 for Empathy in Pearson correlation on the development set.

– In the future, we plan to introduce larger datasets for **model re-training** at earlier stage (e.g. the Masked Language Model), and we plan to consider introducing **conversational context and speaker personality** for better model construction.