Evaluating Significance of Emotion Classification in Emotion-Aware Empathetic Dialogue Systems

Akshen Kadakia
akshenhe@usc.edu

Mrinal Kadam
mrkadam@usc.edu

Swetha Sivakumar
swethasi@usc.edu

Yashwanth Reddy Gondi
gondi@usc.edu

Tanmey Saraiya
tsaraiya@usc.edu

1 Tasks that have been performed

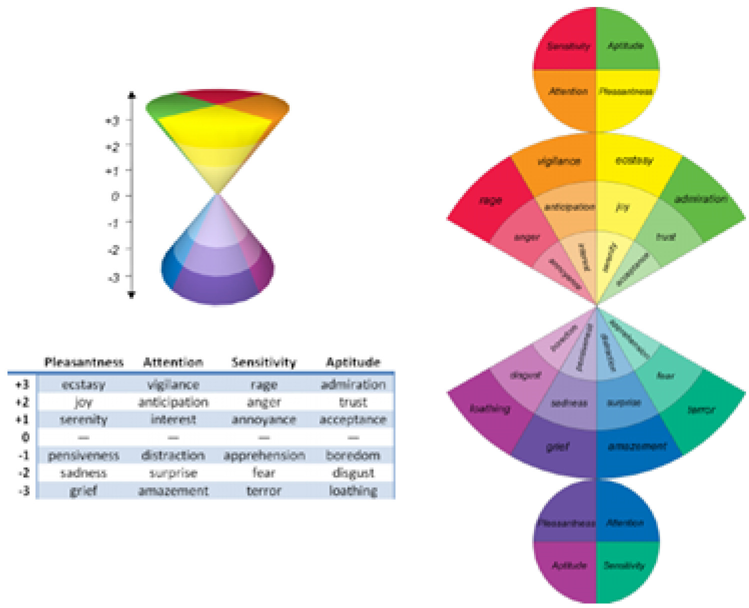
- We have researched about the different possible models that can be used in replacement to the pre-trained emotion classifier using FastText. To train the 'Emotional Awareness' quotient of the dialogue system, the possible candidate architectures are:
  1. **LSTM:** LSTM neural networks can model the temporal characteristics and the context of the data better by learning long term dependencies in the text.
  2. **LSTM with Attention:** Attention can improve the performance of the vanilla LSTM model by extending the traditional encoder-decoder system by paying attention to specific words in the input sequence for each word in the output sequence. It's done by placing different focus on different words and by assigning each word with a score. Then using the softmax scores, the encoder hidden states are aggregated using their weighted sum to get the context vector.
  3. **Transformer based model:** Transformers have proved to be even more effective than LSTMs as they are based on self-attention mechanisms. They are designed to take the whole input sequence at once and therefore significantly allow for more parallelization, thus reaching a new state of the art performance.
- We have finalized the broad categories into which we want to place our 32 fine-grained emotions according to the the hour-glass of emotions as seen in Figure 1.

We want to check whether there has been any significant increase/decrease in our model's performance or not post performing the two tasks mentioned above.

2 Risks and Challenges to address by the project deadline

- **Expensive and Long Training Times:** Training large-capacity models requires heavy computational investment and despite computational resources, the training times are higher. The current work uses Full Transformer Architecture i.e. Encoder and Decoder.
  1. **Retrieval Based Models:** Uses a variety of Transformer-based architectures and BERT-based models.
  2. **Generative Models:**Uses Full Transformer architecture i.e. Both Encoder and Decoder.

  Even to re-achieve the results from the base paper, a major challenge is the requirement of heavy computational resources. And from the extract in Figure 2 from the base paper, we see that despite 8 GPUs the BERT-based models like the PRETRAINED-BERT-R (a retrieval model) model takes 13.5 days to train on 1.7 Billion data points. This reinstates our major challenges, firstly the need for high computational resources and secondly longer training times despite utilization of rich computational resources.
- **Ensuring Domain Consistency:** As per our proposal, we modularly replace the current external predictors with state of the art Emotion-classification models in order to integrate the external predictors which are the Emoprepend and TopicPrepend variations. A major challenge that we expect to face during this replacement process is to maintain the performance. In order to avoid any performance degradation, it is required that the new replacement emotion-classification model that we will introduce be pre-trained



	Pleasantness	Attention	Sensitivity	Aptitude
+3	ecstasy	vigilance	rage	admiration
+2	joy	anticipation	anger	trust
+1	serenity	interest	annoyance	acceptance
0	—	—	—	—
-1	pensiveness	distraction	apprehension	boredom
-2	sadness	surprise	fear	disgust
-3	grief	amazement	terror	loathing

Figure 1 Hourglass of emotions

	Model	Params, resources, train examples	Emp	Rel	Fluent
Retrieval	Pretrained-R	84.3M, 2.5 days, 8GPUs, 1.7B	2.8	3.0	4.1
	Pretrained-ED	same , same, same	3.5	3.6	4.5
	Fine-Tuned	same , + 0.5 hour, 1 GPU, +22.3k	3.8	3.8	4.4
	Pretrained-Bert-R	217M, 13.5 days, 8GPUs , 1.7B	3.1	3.3	4.2
	Pretrained-Bert-ED	same, same, same	3.4	3.5	4.4
	Fine-Tuned-Bert	same, +1hour, 8GPUs, +22.3k	3.7	3.8	4.6
Generative	Pretrained	85.1M, 2 days, 32 GPUs, 1.7B	2.3	2.2	3.9
	Fine-Tuned	same , +1 hour, 1 GPU, +22.3k	3.3	3.3	4.3
	Pretrained-Large	86.2M, 2.5 days, 32 GPUs, 1.7B	2.8	3.0	4.0
	Fine-Tuned-Large	same , +0.5 hour, 1 GPU, +22.3k	3.6	3.6	4.5

Figure 2 Training time and resources of experimental models. Taken from (Rashkin et al. 2019)

on a similar domain with balanced data, which is grounded in emotions and stresses on the empathetic aspects of the dialogue.

3 Plan to mitigate the Risks and address the Challenges

- Computing cost remains to be a major technical challenge since the training time is very high depending on the data set. We plan to address this by applying for AWS compute credits in order to perform training on the cloud.

- Another challenge is consolidating the 32 labels from the data set into 6 labels while making sure that the 6 labels are balanced. To accomplish this, we plan to carefully select them so that the totals are as equal as possible.

References

Rashkin, Hannah et al. (July 2019). *Towards Empathetic Open-domain Conversation Models: A New Benchmark and Dataset*. Florence, Italy: Association for Computational Linguistics, pp. 5370–5381. DOI: 10 . 18653 / v1 / P19 - 1534. URL: https : / / aclanthology.org/P19-1534.