Team ID: 8
Team Name: Binary Squad

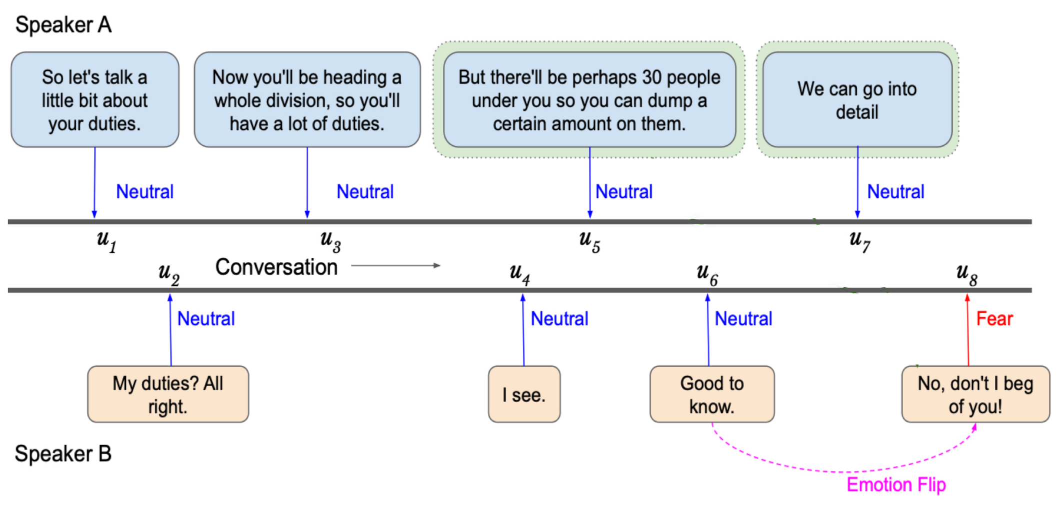
Emotion Discovery and Emotion Flip Detection in Conversation

UNIVERSITÉ Concordia UNIVERSITY

Abstract

- This project focuses on two tasks, Emotion Discovery and Emotion Flip Detection in conversations using the MELD dataset. Emotion discovery (Task A) identifies expressed emotions. We utilized DistilBERT, fine-tuned with techniques to mitigate class imbalance, achieving an accuracy of 50%, alongside insights from F1-score and confusion matrices. This helped in the Task B Emotion Flip Detection, which achieved 99% accuracy, effectively identifying transitions between emotions. The MELD dataset has a few challenges such as class imbalance and emotion overlap, which were addressed through targeted preprocessing and sampling strategies. Our findings highlight the importance of metrics for performance evaluation and finding the potential of fine-tuned transformer models for understanding emotional dynamics in conversations. Future work could explore multi-modal approaches and broader datasets to enhance real-world applicability.

Goal of the Project



The goal of this project is to develop a system for emotion recognition and emotional shift detection in conversations. The system should:

- Accurately classify each utterance into one of the seven emotion categories (joy, anger, sadness, surprise, neutral, disgust, and fear).
- Identify emotional shifts between consecutive utterances to understand dynamic changes in conversations.

Methodology

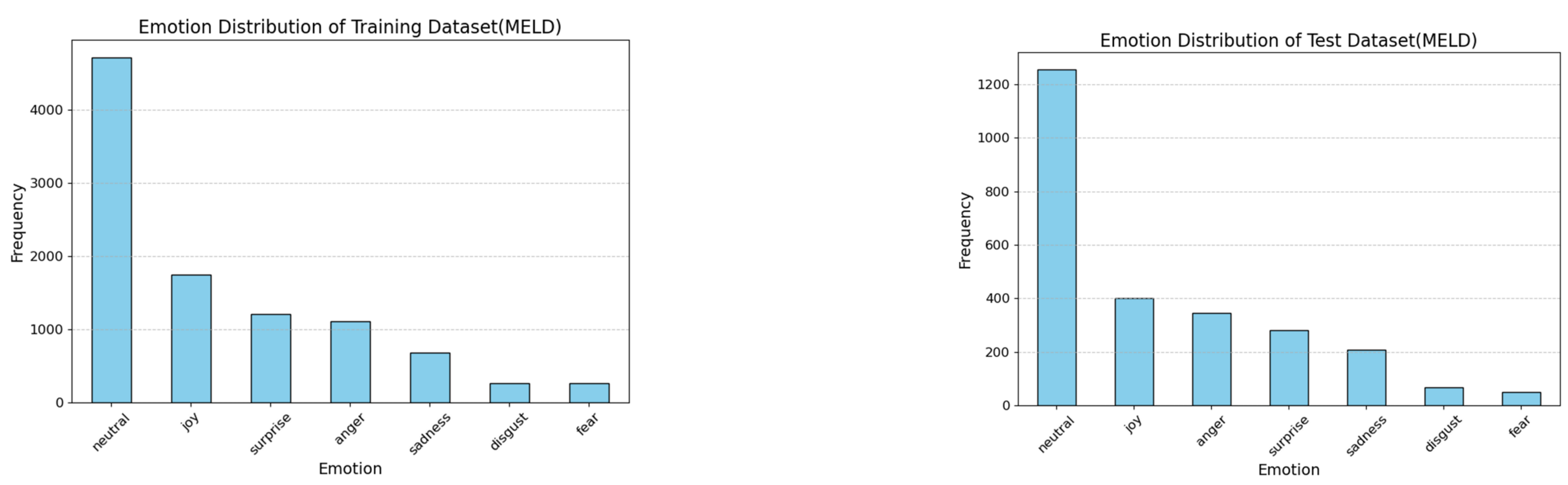


Figure 1. Class Distribution in the Dataset

- **Setup and Data Preprocessing:**
  - Imported essential libraries for model implementation, data handling, and evaluation.
  - Prepared the MELD[1] dataset for training and evaluation through text tokenization using the `DistilBertTokenizer` tokenizer.
- **Model Training:**
  - Implemented and fine-tuned DistilBERT model for emotion classification.
  - Addressed class imbalance using focal loss[2] as an alternative to cross-entropy loss.
  - Optimized the focal loss parameters (gamma) and adjusted the learning rate and batch size for better performance.
- **Evaluation:**
  - Evaluated model performance using the metrics precision, recall, F1-score, and accuracy.
  - Visualized results with confusion matrices, ROC Curves and PR Curves.
- **Task B - Emotion Flip Detection:**
  - Extended the fine-tuned emotion classification model to detect emotional shifts between utterances.

Results

Dataset	MELD (Random Sample Of 1500)
Model	DistilBERT
Loss Function	FocalLoss
Learning Rate	1.00E-05
Max Length of Embedding	95
Batch Size	32 (Optimal)

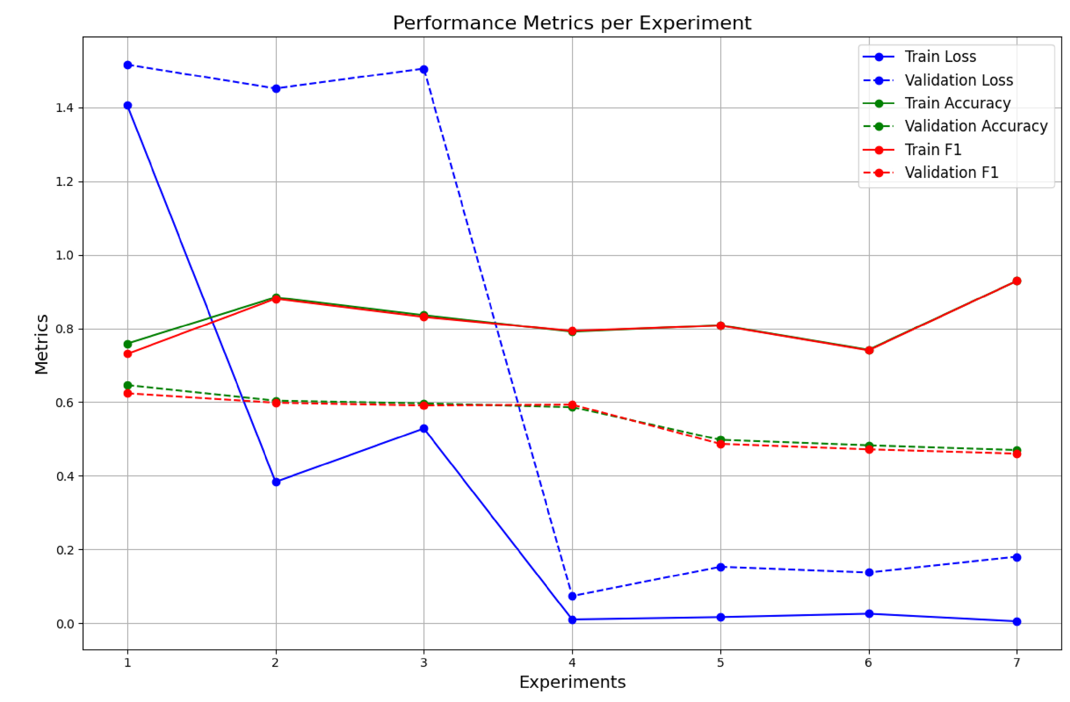


Figure 2. Optimal Exp. Config | Visualizing Metrics of Experiments

Metrics	Train Loss	Train Acc	Train F1	Val Loss	Val Acc	Val F1
First Exp(Task A)	1.4057	0.7591	0.7311	1.5162	0.6460	0.6239
Optimal Exp(Task A)	0.0165	0.8089	0.8080	0.1532	0.4977	0.4866
Task B	-	-	-	-	0.99	0.75

Figure 3. Metrics Comparison of Initial and Optimal Model Config.

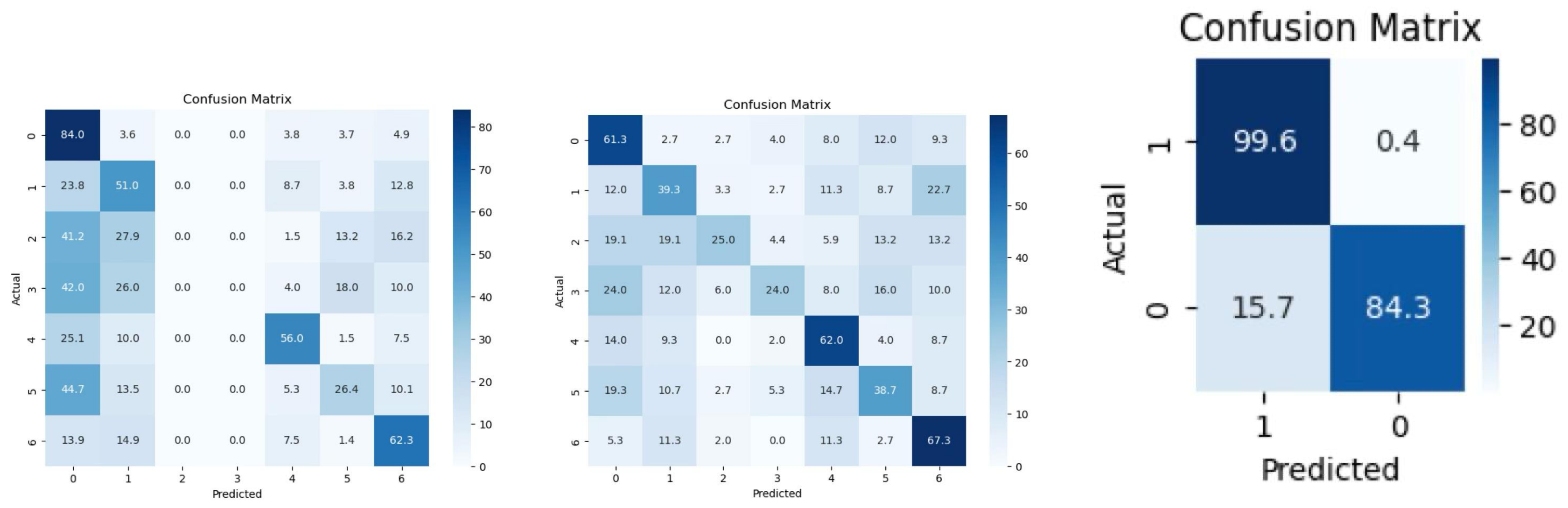


Figure 4. Confusion Matrix of Task A First Model | Optimal Model || Task B

Analysis

- Initially, the model struggled to predict disgust and fear due to class imbalance. One of the solutions attempted was to add more data by combining the MELD and DailyDialog datasets. However, because the conversational styles of the two datasets were different, this resulted in a negative effect on model performance.
- Switching from BERT to DistilBERT improved predictions for 'disgust' and 'fear' because DistilBERT, being a smaller and more efficient model, it reduced overfitting and allowed the model to better generalize on the imbalanced dataset.
- Focal Loss significantly improved performance for all classes by focusing more on minority classes which had fewer instances.
- The model achieved optimal performance during Experiment 5, as seen in the metrics plot.
- Further experiments saw declining performance, due to overfitting or suboptimal hyperparameter changes.

Limitations and Future Work

- Limitations: Computational overhead, limited data, and class imbalance affected models ability to effectively generalize.
- Future Work: To create more conversational datasets from diverse sources such as sitcoms. Additionally, develop a model capable of providing reasoning for emotional shifts, enhancing and understanding of emotion flips in conversations.

References

[1] S. Kumar, M. S. Akhtar, and E. T., "Semeval 2024 – task 10: Emotion discovery and reasoning its flip in conversation," *arXiv preprint arXiv:2402.18944*, 2024.

[2] A. Yadav, "Implementing focal loss in pytorch for class imbalance," *Medium.com*, 2024.