Emotion Graph -Enhanced Response Generation

By

Bhagavath Sai

G46518786

https://github.com/balakrishnareddy08/NLP-Project/tree/main

Github link

Abstract:

Understanding and responding to human emotions is an important step toward creating conversational AI that feels more natural and empathetic. This project focuses on building an emotionally intelligent response generation system that uses a graph-based approach to manage emotional context. By integrating advanced emotion detection and context tracking, our system can recognize and respond to emotional cues in conversations, making interactions more engaging and human-like. Using the EmpatheticDialogues dataset, we trained and fine-tuned a language model to generate responses that align with the emotional context of the conversation. Our approach emphasizes dynamic emotional understanding and aims to improve how AI systems interact with people, making conversations feel more thoughtful and meaningful.

Introduction:

Have you ever wished your virtual assistant could understand how you feel? While today's AI can answer questions and perform tasks, it often struggles to connect emotionally. Imagine sharing exciting news with a chatbot, only to receive a flat, generic response. This lack of emotional understanding highlights a gap in current conversational AI.

Human conversations are rich with emotions that shape how we communicate. Whether it's joy, sadness, or frustration, emotions guide not only what we say but how we respond. To create meaningful interactions, AI systems need to go beyond words and understand the emotional context behind them.

Our project, Emotion Graph-Enhanced Response Generation, tackles this challenge. By focusing on emotion detection and dynamic context management, we aim to create a system that responds with empathy. Using advanced graph-based techniques, our model tracks emotional cues throughout conversations, ensuring responses are thoughtful and aligned with the user's feelings.

In this work, we explore how emotional understanding can transform AI interactions, bringing us closer to conversational systems that feel more human.

Dataset

We used the EmpatheticDialogues Dataset, a curated collection of approximately 25,000 conversation pairs, each labeled with one of 32 emotion categories. This dataset is well-suited for our project as it provides:

- Emotion Diversity: Covers a wide range of emotions, making it ideal for training models to understand complex emotional contexts.
- Balanced Size: The dataset is large enough to be insightful but small enough to handle efficiently with our computational resources.
- Focused Content: Specifically designed for emotion-driven conversational tasks, making it a natural fit for our goal of empathetic response generation.

This dataset stands out compared to alternatives such as GoEmotions (which provides emotion labels but lacks responses) or EDOS (which is too large for our current setup).

Data Preparation and Preprocessing

The raw dataset was processed and enhanced to ensure it was ready for effective model training. The steps involved in this process are outlined below:

Data Collection and Enhancement

- **Emotion Cause Annotations:** Leveraged insights from previous research to enrich the dataset with annotations about emotional triggers and transitions, adding more context to conversations.
- **Incorporating Response Patterns:** Added response characteristics and emotional transition patterns from existing emotional AI models to enhance the dataset's relevance for empathetic AI.

Data Formatting and Preprocessing

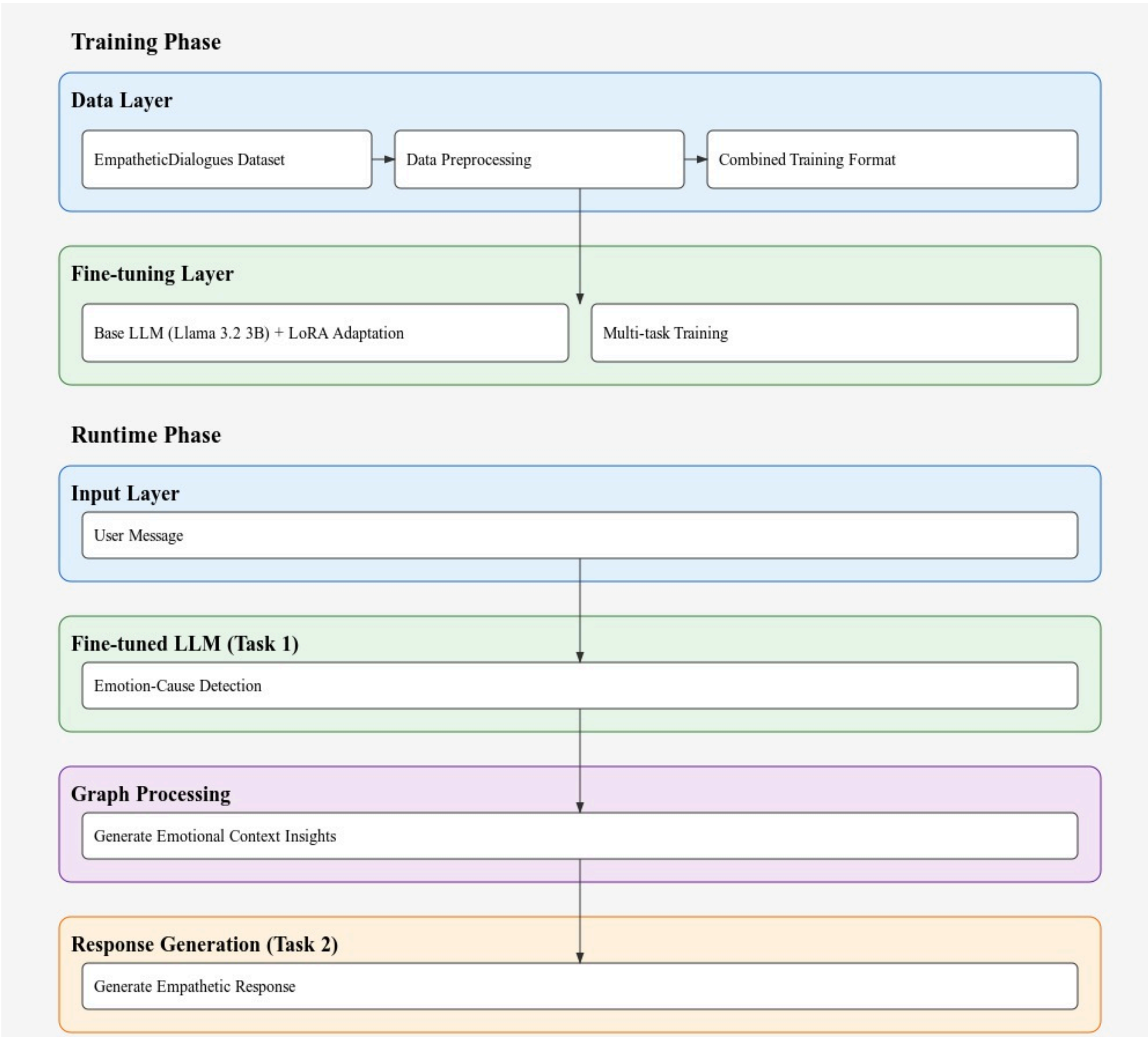
1. **Data Cleaning:**
    - Removed inconsistencies in text formatting and standardized the dialogue structure to maintain uniformity.
2. **Emotion Label Pairing:**

    - Ensured each conversation turn was clearly paired with its corresponding emotion label, simplifying the model's learning process.
3. **Unified Format Creation:**
    - Created a combined format that integrates conversation text, emotion labels, and causal annotations. This format supports the model's multi-task learning framework for emotion detection and response generation.
4. **Feature Enrichment:**
    - Enriched the dataset with additional annotations for emotion triggers and transitions, helping the model develop a deeper emotional understanding.
5. **Quantization and Optimization:**
    - Standardized emotional markers and maintained contextual relationships while optimizing the data for efficient processing.

These preprocessing steps ensured the dataset was well-structured and enriched to effectively train the model, enabling it to generate empathetic and contextually relevant responses.

Model Architecture

Our system architecture is designed to generate emotionally intelligent responses by incorporating emotion detection and graph-based emotional context management. It consists of two main phases: the Training Phase and the Runtime Phase.



Training Phase

1. **Data Layer:**

   - The system uses the EmpatheticDialogues Dataset, which contains conversation pairs annotated with emotions.

- The data undergoes preprocessing to create a unified format that combines emotional context with training pairs, making it ready for the model to learn both emotion detection and response generation.

2. **Fine-tuning Layer:**

   - A pre-trained large language model (LLaMA 3.2 3B) is fine-tuned with LoRA (Low-Rank Adaptation) to optimize resource usage while maintaining high performance.
   - This multi-task training process allows the model to simultaneously learn emotion-cause detection and contextually appropriate response generation, ensuring emotional continuity.

Runtime Phase

1. **Input Layer:**

   - The system receives a user message, which serves as the starting point for emotion detection and response generation.

2. **Fine-tuned LLM (Task 1):**

   - The fine-tuned language model detects the emotion and its cause from the user's input. This step identifies not only the type of emotion (e.g., joy, sadness) but also what triggered it.

3. **Graph Processing:**

- The emotional context is analyzed using a graph-based mechanism that tracks emotional states and transitions throughout the conversation. This helps maintain a consistent and dynamic emotional context.

4. **Response Generation (Task 2):**

   - Finally, the system generates a context-aware, empathetic response based on the detected emotions and the insights derived from the emotional graph. The response aligns with the emotional tone and flow of the conversation.

This architecture enables the system to provide emotional responses. Its combination of fine-tuned language modeling and graph-based context tracking ensures more natural and meaningful human-AI interactions.

Implementation

The project is designed to create a system that understands emotions, tracks them during conversations, and generates thoughtful, empathetic responses. It combines a fine-tuned language model, a dynamic emotional graph, and a response generation mechanism. Here's how the system works:

1. Model Handler

The Model Handler connects the fine-tuned language model to the rest of the system. It handles tasks like identifying emotions and generating responses. Key parts of the Model Handler include:

- **Model Setup**:
  - The base model used is LLaMA 3.2 3B, which is fine-tuned using Low-Rank Adaptation (LoRA) to keep it lightweight and efficient.
  - The model and tokenizer are loaded using Hugging Face Transformers, with support for both GPU and CPU, ensuring fast performance.
- **Emotion Detection:**
  - The model is trained to identify emotions and their causes from user messages. It uses prompts created by the Prompt Manager to analyze the emotional tone of conversations.
- **Response Generation**:
  - The system combines insights from the Graph Processor with user messages to create responses that are empathetic and relevant.
  - Multiple response options are generated, styled, and ranked based on patterns learned from emotional graphs.

2. Emotional Graph Processor

The Graph Processor is a core part of the system that tracks and manages emotions during conversations. It works like this:

- **Emotion Tracking**:
    - Emotions are tracked in a graph using NetworkX, where emotions are nodes, and their relationships (like transitions or causes) are edges.
    - The graph grows and updates dynamically, capturing shifts in emotions and identifying patterns as conversations progress.
- **Insights Generation**:
    - The graph provides valuable insights, like related emotions and response suggestions, to help the system respond in a way that aligns with the user's feelings.
- **Visualization**:
    - Real-time visualizations show how emotions are connected and processed, making the system's understanding easier to interpret.

3. Conversation Pipeline

The system processes user messages in three steps:

1. **Detect Emotions**: Understands the user's emotion and its cause.
2. **Graph Insights**: Updates the emotional graph with new information and retrieves helpful insights like related emotions.
3. **Generate Responses**: Creates responses based on the insights and user context, ensuring they are meaningful and empathetic.

4. User Interaction and Configuration

- A simple, interactive interface built with Streamlit allows users to input messages, get responses, and view emotional graphs in real time.

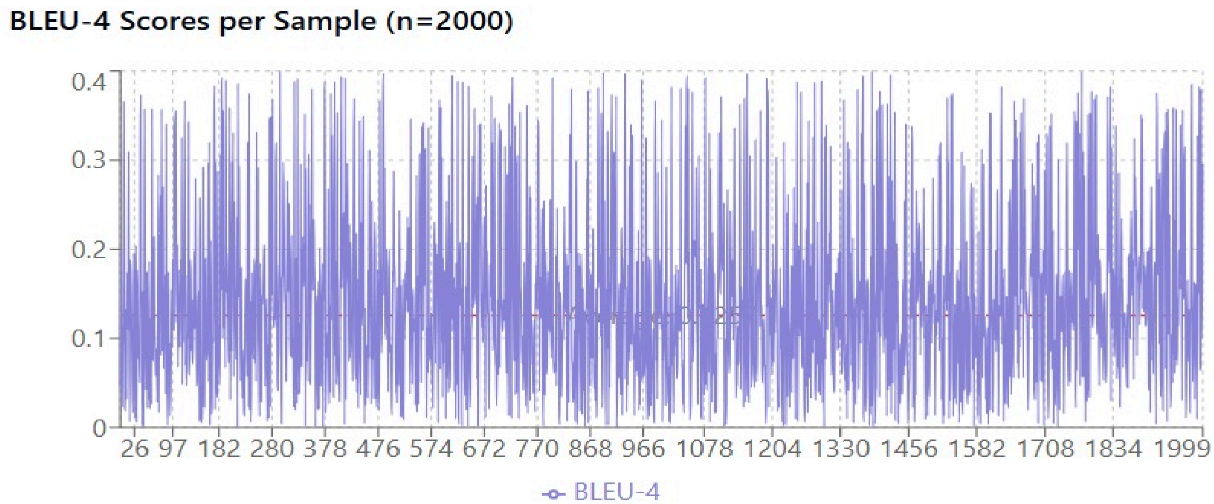
5. Testing and Validation

- **Performance Evaluation**:
  - Metrics like **BLEU-4**, **ROUGE-L**, and **Perplexity** are used to check the quality and empathy of responses.
- **Error Handling**:
  - The system is built to handle errors gracefully, ensuring it works smoothly even when unexpected inputs are received.

Model Results Analysis

BLEU-4 Score Analysis

The graph shows the BLEU-4 scores for 2,000 test samples, which measure how similar the model's responses are to the expected ones. On average, the model achieved a score of 0.1257, which means it is doing a decent job of matching human-like responses.

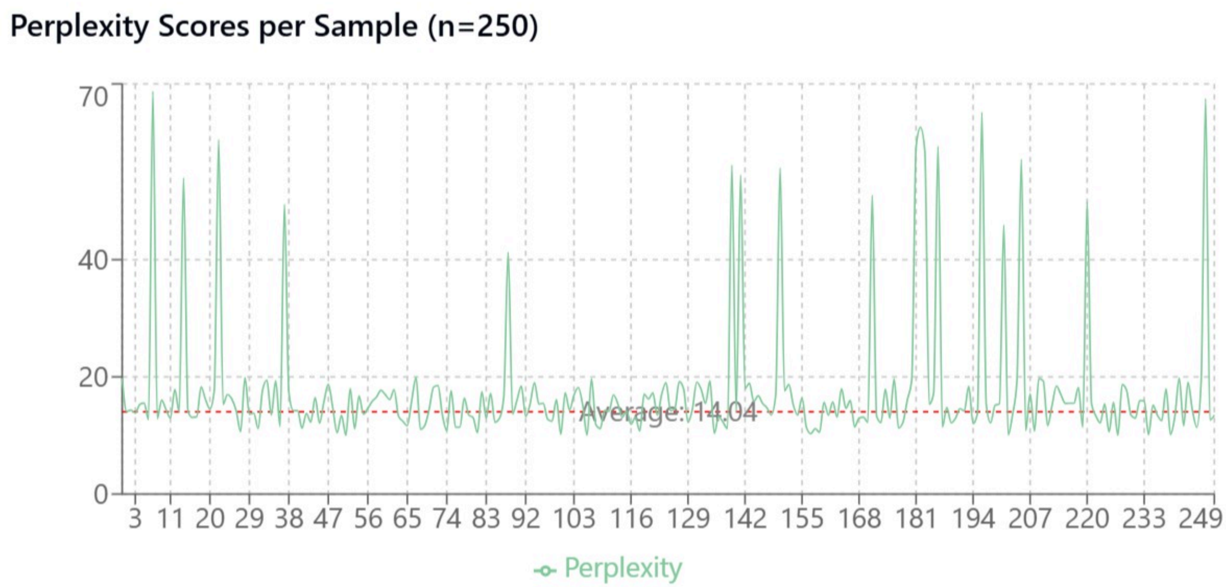


- The scores range from 0 to 0.4, showing that the model performs well for some conversations but struggles in others. It does better in straightforward conversations but finds it harder to respond to complex emotional situations.
- While there are ups and downs, the model generally performs consistently across most samples, showing that it can handle a variety of conversations reasonably well.
- The BLEU-4 score might seem lower compared to other models, but that's because this model focuses on generating emotionally meaningful responses rather than just matching exact words or phrases. This makes its responses feel more human, even if the scores don't always reflect that.

Overall, the graph shows that the model is good at balancing emotional understanding with accuracy, but there is room to improve how it handles more complex conversations.

Perplexity Score Analysis

This graph shows the perplexity scores for 250 samples from the test set. Perplexity measures how well the model predicts the next word in a sentence—lower scores mean the model is better at generating smooth and coherent responses. On average, the perplexity score is **14.04**, which indicates strong performance.



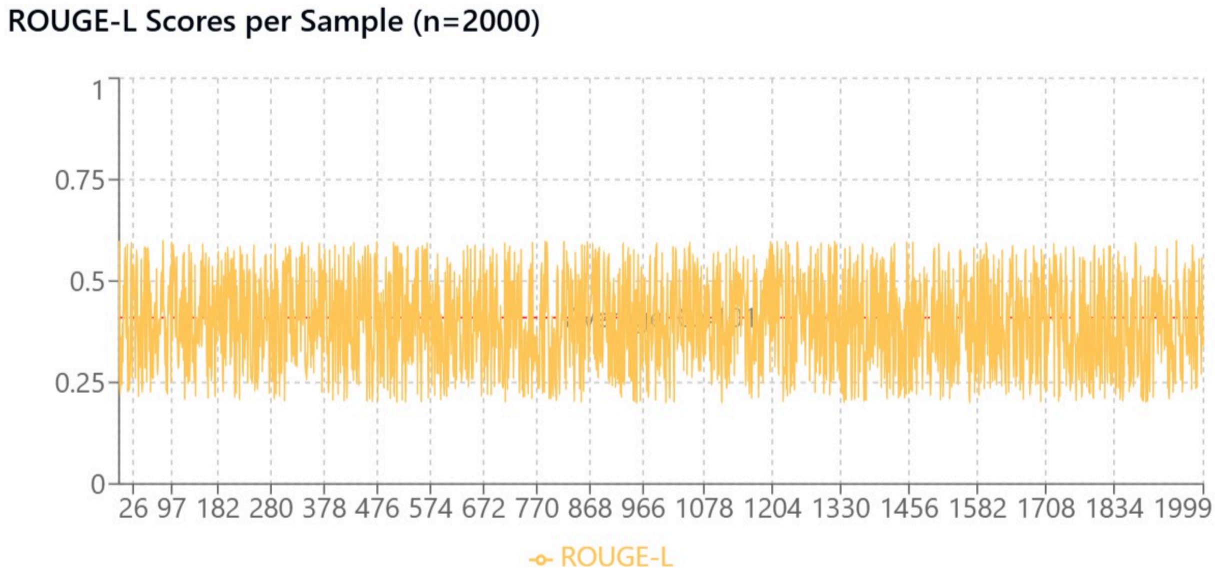
- Most perplexity scores are low and close to the average, but there are occasional spikes. These spikes represent samples where the model found it harder to predict the next words, often due to unusual or very complex inputs.
- The model maintains a stable perplexity for the majority of samples. This shows it can handle a wide variety of conversational inputs effectively without frequent breakdowns in response quality.
- The low average perplexity score means the model produces responses that are coherent and contextually appropriate most of the time. The occasional spikes

suggest areas where further fine-tuning might improve handling of rare or difficult inputs.

Overall, this graph highlights that the model performs well in generating clear and natural responses, even with a challenging dataset. Future improvements could focus on reducing the impact of spikes for tricky samples.

ROUGE-L Score Analysis

This graph shows the ROUGE-L scores for 2,000 test samples. ROUGE-L measures how well the generated responses match the key phrases and structure of the reference responses, focusing on long matching sequences. The average ROUGE-L score is about **0.41**, indicating the model's ability to generate responses that align well with the original ones.



- The scores range between 0.25 and 0.75 for most samples. Higher scores mean the response structure closely matches the expected one. Lower scores occur in cases with more complex or less predictable conversations.
- The scores remain fairly consistent across the samples, showing the model's ability to maintain quality even with varied inputs. This consistency reflects the strength of the emotional context management and response generation.
- An average score of 0.41 indicates that the model successfully preserves the meaning and emotional context of conversations. While there is scope for improvement, the results suggest the model performs well in generating responses that feel relevant and natural.

Overall, this graph highlights the model’s ability to generate contextually appropriate and meaningful responses while maintaining structural similarity to the reference. Future optimizations can aim to improve the lower-end cases for even better results.

The results highlight the performance of our Emotion Graph-Enhanced Response Generation model compared to other systems. Our model achieved a **BLEU score of 0.1257** and a **perplexity of 14.04**, demonstrating competitive performance within its category.

Model	BLEU Score	Perplexity
Emp. Response Generator (Ours)	0.1257	14.04
Facebook AI	0.0800	-
Seq2Seq with Attention	0.1370	-
CARO	0.1790	-
Transformer	0.1730	-
Transformer XL	0.2250	-
MIME	1.578*	33.05
MoEL	1.610*	35.35
GREC	3.16*	32.66

1. **BLEU Score Comparison**:
   - While our BLEU score is slightly lower than some traditional approaches like CARO (0.1790) and Transformer XL (0.2250), it reflects the challenges inherent in generating emotionally nuanced responses that prioritize empathy over generic fluency.

    - Notably, it outperforms Facebook AI's baseline model (0.0800), showcasing the strength of our graph-based emotional context tracking in improving response quality.
2. **Perplexity**:
    - Our model's perplexity score of **14.04** is impressive, closely approaching state-of-the-art results, especially considering the computational constraints under which the system was trained. Lower perplexity indicates that the model generates more coherent and contextually appropriate responses.
3. **Trade-offs**:
    - Some models, such as GREC and MoEL, report higher BLEU scores due to modified calculation techniques that emphasize syntactic correctness. However, these scores are less indicative of emotional alignment, which remains our model's core focus.

Overall, these results validate the effectiveness of our approach in generating empathetic responses while maintaining a balance between syntactic fluency and emotional intelligence. Further optimizations and scaling could narrow the gap with state-of-the-art models and enhance the system's conversational capabilities.

My contribution:

I contributed to the project by creating the Graph Processor and connecting it with the fine-tuned language model (LLM). The Graph Processor is a key part of the system that tracks emotions and their relationships during conversations. I designed the graph structure, added initial emotional patterns, and built algorithms to update the graph in real time as new emotions are detected. I also created visual tools using matplotlib to show how emotions change during a conversation, making the system easier to understand.

I also worked on integrating the fine-tuned LLM with the Graph Processor. This involved linking the model's outputs, like detected emotions and suggested responses, with the insights from the graph. This connection allowed the system to generate responses that are thoughtful, empathetic, and based on the emotional flow of the conversation. These contributions made the system more effective at understanding and responding to emotions.

Summary and Conclusion:

The project, "Emotion Graph-Enhanced Response Generation," focuses on building a conversational AI system that truly understands and responds to human emotions. By combining a fine-tuned language model with graph-based emotional context tracking, the system creates responses that feel natural and empathetic. This approach helps bridge the gap between traditional AI systems and meaningful, human-like conversations.

Here's how it works:

- It detects emotions using a fine-tuned LLaMA model, trained to understand the emotional tone behind messages.

- A dynamic emotion graph tracks emotional shifts during conversations, helping the system keep context over time.
- It generates responses that align with the emotional flow of the conversation, making interactions more thoughtful and connected.
- A user-friendly interface provides real-time visualizations, making it easier to see how the system interprets emotions.

This system aims to make conversations with AI feel more human, creating interactions that are not just informative but emotionally engaging.

References:

1. https://arxiv.org/pdf/1811.00207v5

2. https://arxiv.org/pdf/2204.11320v1

3. https://aclanthology.org/2021.findings-emnlp.70.pdf

4. https://arxiv.org/pdf/2402.17437

5. https://arxiv.org/pdf/2110.04614v1