THERAPY CHATBOT

Team:
Bright Ofori
Siri Gangadharaiah

Date:
12/06/2024

Project Overview and Objectives

This project focuses on creating a chatbot for therapy which is mainly concerned with a person’s mental health.

Experimented with different datasets and with various models to check the quality of the chatbot’s response.

- Focuses on a person’s emotions and sentiments.
- Being able to understand the context and have a therapeutic conversation with the person.
- Experiments focus on:
    - Classifying suicidal tendencies
    - Conversational-based experiments

Suicidal Classification

- **Data Source:** Chose a dataset from Kaggle to finetune the model.
- **Objective**: Classify if the user's case is suicidal or not.
- **Methodology**:
  - Text classification model using NLP's pre-trained model BERT
  - Dataset: Annotated data of conversations related to mental health
  - Approach: model training, validation and then fine-tuned the model using the selected dataset
- **Outcome**: Classifies if the case is suicidal or non-suicidal based on the person's input

Results

```
              precision    recall  f1-score   support

 non-suicide       0.98      0.98      0.98     11604
     suicide       0.98      0.98      0.98     11604

    accuracy                           0.98     23208
   macro avg       0.98      0.98      0.98     23208
weighted avg       0.98      0.98      0.98     23208
```

```python
sample_text_1 = "I feel hopeless and don't want to live anymore."
sample_text_2 = "I had a great day today and feel very happy!"

print("Prediction for Sample 1:", predict_text(sample_text_1))  # Expected: "suicide"
print("Prediction for Sample 2:", predict_text(sample_text_2))  # Expected: "non-suicide"
```

```
Prediction for Sample 1: suicide
Prediction for Sample 2: non-suicide
```

Chatbot – For Therapeutic Conversation

Experimented with 2 models with 2 datasets

Datasets: The one dataset contains conversations between users and experienced psychologists related to mental health topics

The other dataset is labeled, with each question corresponding to a specific user (patient) and the answer provided by a therapist.

Models Used:

- **Model 1:** Meta LLaMA 3 (3B parameters)
- **Model 2:** Meta LLaMA 3 (8B parameters)

Model 1: Meta LLaMA 3 (3B parameters)

- **Fine-Tuning Approach:** The model was fine-tuned on our custom dataset.
- Trained using 3 epochs
- **Challenges and solutions :**
  - The 3B parameter size posed significant computational demands.
  - To address this, the model was loaded in an 8-bit quantized format with float16 precision to reduce memory usage.
  - Performance-efficient fine-tuning (PEFT) techniques were employed to optimize training on limited resources.
  - Gradient clipping was applied to manage memory consumption effectively and ensure stable training.
- **Results:** Performance was satisfactory but highlighted potential areas for improvement, particularly in terms of generalization and accuracy on complex tasks.

Model 2: Meta LLaMA 3 (8B parameters)

- **Fine-Tuning Approach:** The larger 8B parameter model was fine-tuned on the same custom dataset.
- Trained using 3 epochs
- **Challenges and solutions :**
  - The larger model significantly increased computational complexity and resource demands.
  - Similar to Experiment 1, the model was loaded in an 8-bit quantized format with float16 precision.
  - Advanced PEFT techniques were employed to handle the scale of the model.
  - Gradient clipping was used to ensure memory optimization and prevent gradient explosion.
- **Results:**
  - The 8B parameter model demonstrated markedly better performance compared to the 3B model.
  - Improvements were observed in:
    - Accuracy on the custom dataset
    - Ability to handle nuanced and complex queries

Evaluation Metrics


```
{
  "generation_metrics": {
    "perplexity": 62.72367858886719,
    "generation_length": 9.0
  },
  "text_quality_metrics": {
    "unique_words_ratio": 0.8666666666666667,
    "avg_word_length": 3.8333333333333333
  },
  "comparison_metrics": {
    "exact_match": 0.0,
    "similarity_score": 0.22376543209876543
  }
}
```

Recommendations for future work:

- **Fine-tuning**: Augment the training data with examples of nuanced or mixed-sentiment messages to reduce false positives.
- **Threshold Adjustment**: Apply a confidence threshold (e.g., 0.9 for LABEL_1) to minimize the overclassification of ambiguous messages as suicidal.
- **Hardware Utilization**: Ensure GPU utilization by setting the device parameter in the pipeline to speed up inference and potentially enhance real-time applications.
- **Explainability**: Implement interpretability techniques (e.g., SHAP or LIME) to better understand why the model flags certain messages.

Conclusion and Next Steps

Key Observations:

- Larger models (8B parameters) have a clear advantage in terms of performance and accuracy, though they require more computational resources.
- Efficient techniques like quantization, PEFT, and gradient clipping are essential to make training on large models feasible.

Next Steps:

- Conduct further fine-tuning with additional epochs to explore potential gains.
- Investigate parameter-efficient inference techniques to optimize real-world deployment.
- Explore additional optimizations like mixed-precision training and advanced PEFT methods for scaling to even larger models.

Thank You!