CSE556: Natural Language Processing 2025
Project Mid-Evaluation
Mental Health Counselling Summarization

Himanshu Raj 2022216
Ishita 2022224
Ritika Thakur 2022408

Abstract

This project focuses on developing an NLP-based system for generating concise and informative summaries of mental health counselling conversations. We aim to capture the key insights of therapeutic dialogues while preserving their contextual nuances.

1 Introduction

In recent years, increased awareness of mental health has led to a growing demand for effective counselling services. Mental health counselling dialogues are inherently complex, comprising nuanced exchanges that include emotional expressions, therapeutic interventions, and reflective insights. These characteristics make it challenging to generate concise summaries that capture the key insights while preserving the context and subtleties of the conversations.

We aim for the task of counselling conversation summarization. The motivation behind this work stems from the need to assist mental health professionals in quickly grasping the core content of lengthy dialogues, thereby reducing the cognitive load and improving the efficiency of documentation and analysis.

2 Related Work

Dialogue summarization has evolved to tackle the unique structure and complexity of conversational data. Unlike news or narrative text, dialogues often involve frequent topic shifts, role-specific utterances, and non-informative fillers, making summarization more challenging.

Recent work has begun to address these challenges in domain-specific contexts. Srivastava et al. (Srivastava et al., 2022) propose a method for counseling dialogue summarization that uses mental health knowledge to filter out irrelevant content and highlight therapeutic elements. Their approach focuses on identifying clinically meaningful utterances, improving the relevance and clarity of generated summaries.

Other studies have incorporated discourse structure, speaker roles, and contextual embeddings to improve summary quality in dialogue settings. However, most methods—including those using large language models like T5—treat all utterances with equal weight, often overlooking the emotional nuances that are especially important in counseling and mental health contexts.

Our work builds on this foundation by exploring the integration of emotional salience into the summarization process, addressing a gap in how affective cues are used to guide both content selection and summary generation.

3 Methodology

We proposed to use the emotional information of the utterances in the dataset to improve summarization and the quality of the generated summaries. We are using Pegasus-Xsum and T5-base variant for our experiments.

The first method employs a two-stage process to enhance summarization. Initially, a pre-trained emotion recognition model (Raw, 2020) analyzes each utterance in the dataset, assigning corresponding emotion tags. These tags are then integrated with the textual data, providing explicit emotional context for each utterance. Subsequently, a pre-trained summarization model, such as T5 or Pegasus, is fine-tuned on this emotion-annotated dataset. The model learns to associate emotional cues with the source text during this process. By conditioning the summarization process on the detected emotions, this approach aims to generate summaries that are not only factually accurate but also more

sensitive to the emotional nuances of the original content, potentially leading to improved relevance, coherence, and a more human-like quality in the generated summaries.

The second method involves fine-tuning a pre-trained summarization model on larger datasets to improve its capabilities and expressiveness so that it performs better when fine-tuned on our intended dataset (MEMO). We fine-tune T5 on GoEmotions so that the model understands the emotional context in the text and DialogSum so that the model understands how to summarize a conversational dialogue between two people. This multi-dataset fine-tuning aims to equip the model with a robust understanding of how emotions manifest in conversational settings and how dialogues are typically summarized. To facilitate efficient training on these large datasets, Low-Rank Adaptation (LoRA) (Face, 2023) is employed, which significantly reduces the number of trainable parameters. Further, this model is fine-tuned on the MEMO dataset. To maintain training efficiency and potentially adapt to the nuances of the MEMO dataset, another LoRA layer is added during this final fine-tuning stage. This layered fine-tuning approach, utilizing larger emotional datasets and efficient LoRA training, aims to produce a summarization model that is particularly adept at capturing and conveying the emotional subtext within the dataset's summaries.

4 Dataset, Experimental Setup, and Results

4.1 Datasets

All the datasets used are exclusively presented in the English Language.

MEMO (MEMO) dataset contains 12,900 utterances from 212 counselling conversations between the therapist and the patient. The train, validation and test sets contain 152, 21 and 39 conversations, respectively. The conversations have utterances, summaries, primary topics and secondary topics. Each utterance has a subtopic (Sub Topic), speaker (Type), utterance id (ID), and an emotion (Emotion) field associated with it. The subtopic field has psychotherapy elements like symptom and history, patient discovery, reflecting, inactive, etc., which can help identify discussion fillers. The emotion field is empty for utterances across all 212 samples.

The Google AI **GoEmotions** dataset consists of 58,009 comments from Reddit users with labels of their emotional colouring. It is designed to train neural networks to perform deep analysis of the tonality of texts. The categories of emotions were identified by Google together with psychologists and include 12 positive, 11 negative, 4 ambiguous emotions, and 1 neutral, making the dataset suitable for solving tasks requiring subtle differentiation between different emotions. The emotion categories are admiration, amusement, anger, annoyance, approval, caring, confusion, curiosity, desire, disappointment, disapproval, disgust, embarrassment, excitement, fear, gratitude, grief, joy, love, nervousness, optimism, pride, realization, relief, remorse, sadness, surprise and neutral. The train, validation and test sets contain 43410, 5426 and 5427 samples, respectively.

DialogSum is a large-scale dialogue summarization dataset consisting of 13,460 dialogues with corresponding manually labelled summaries and topics. The train, validation and test sets contain 12460, 500 and 1500 samples, respectively. Each entry in this dataset offers insights into a wide range of conversational scenarios, capturing interactions among individuals engaged in various everyday life discussions. The dialogues encompass a diverse spectrum of topics, covering areas such as schooling, work, medication, shopping, leisure, travel, and more. These conversations unfold in different real-life settings, featuring exchanges between friends, colleagues, customers, and service providers.

4.2 Experimental Setup

We leveraged the resources and collaborative environment provided by the Kaggle platform for all our experiments. The platform provides a docker container for running Jupyter Notebooks that comes with standard libraries and packages widely used for Machine Learning and Deep Learning tasks.
We used Hugging Face transformers and dataset libraries to access public models and datasets. Pre-trained models were loaded directly from the Hugging Face Model Hub, and fine-tuning was performed using the library's training functionalities.

4.3 Results

The final trained models can be found in this folder.

Model	BLEU4 score	BERT score
T5-baseline	**12.20**	79.41
Pegasus-baseline	11.39	79.81
T5-emo	0.30	79.71
Pegasus-emo	5.57	**86.90**
T5-LoRA	0.33	79.78

```
Original summary: The patient's emotional inventory of stress, worry and anxiety is ra
ted three by the patient. The patient feels slightly better than before. The therapist
suggests two tasks. One is to give a self talk in the morning when the automatic clock
comes up, and next is to note down the automatic thoughts like feelings of anxiety whe
n it occurs in the day. The therapist suggests to maintain a journal so that they can
work on the patient's adaptive responses. The therapist assures to set up an appointme
nt for next week and assures patient of their strong ability to overcome this.

Generated summary: The patient is asked to take a emotional inventory of their stress
and worry and anxiety. The patient stated that they are less anxious now than they wer
e before but they are still suffering. The therapist requests the patient to apply a s
elf talk in the morning to practice what they are doing. The patient did not find the
automatic thoughts to be very helpful. So they are trying to cultivate an awareness in
the day when they are having these thoughts. The patient identifies if they are going
to do something unpleasant, they might feel the need to do something drastic. The pati
ent further adds that they don't like the feeling of
```

Figure 1: Pegasus summary with emotional aspect

5 Observations and Discussion

- **Emotion Integration Challenges**: Incorporating emotion tags into pre-trained models like T5 and Pegasus presented challenges. Notably, the T5-emo model exhibited a significant drop in BLEU4 scores compared to its baseline, suggesting that naive integration of emotional cues can disrupt the model's summarization capabilities. This underscores the necessity for more sophisticated methods to embed emotional context effectively.

- **Performance of Pegasus-emo**: Contrastingly, the Pegasus-emo model demonstrated a substantial improvement in BERTScore, indicating enhanced semantic alignment with reference summaries. This suggests that certain architectures may be more amenable to emotion-aware enhancements, or that the integration method used with Pegasus was more effective.

- **Limitations of LoRA Fine-Tuning**: The T5-LoRA model did not exhibit significant improvements over the baseline. This may point to limitations in the LoRA fine-tuning approach for this specific application or indicate that further optimization is required to realize its potential benefits.

- **Data Annotation Gaps**: The absence of pre-annotated emotion labels in the MEMO dataset necessitated the use of external emotion recognition models. This additional preprocessing step introduces potential inaccuracies and highlights the importance of high-quality, emotion-annotated datasets for training and evaluation.

- **Alignment with Existing Research**: Our findings resonate with recent studies emphasizing the importance of emotional context in counseling summarization. For instance, Srivastava et al. introduced ConSum, a model that filters utterances based on mental health knowledge to enhance summary relevance (Srivastava et al., 2022). Similarly, the EmpRes model integrates sentiment guidance and commonsense reasoning to generate empathetic responses in virtual mental health assistants (**?**). These approaches underscore the growing recognition of emotional nuances in therapeutic dialogue summarization.

6 Conclusion and Future Work

In this project, we investigated the integration of emotional context into the summarization of mental health counseling dialogues. Our experiments revealed that while emotion-aware models like Pegasus-emo can enhance semantic alignment, challenges remain in effectively incorporating emotional cues without compromising summary coherence.

Future Work:

- **Advanced Emotion Integration Techniques**: Explore more sophisticated methods for embedding emotional context, such as attention mechanisms that prioritize emotionally salient utterances or multi-task learning frameworks that jointly model summarization and emotion recognition.

- **Dataset Enhancement**: Collaborate with mental health professionals to annotate existing datasets with emotion labels, ensuring higher quality training data and more accurate evaluations.

- **Model Architecture Exploration**: Investigate the efficacy of other pre-trained models and architectures, including those specifically designed for empathetic response generation, to determine their suitability for emotion-aware summarization tasks.

- **Real-World Application and Evaluation**: Develop a prototype application that utilizes the emotion-aware summarization model, and conduct user studies with mental health practitioners to assess its practical utility and gather feedback for further refinement.
- **Ethical Considerations**: Address ethical concerns related to the use of AI in mental health contexts, ensuring that generated summaries maintain confidentiality, avoid bias, and support the therapeutic process without unintended consequences.

By pursuing these avenues, we aim to contribute to the development of tools that assist mental health professionals in efficiently summarizing counseling sessions, ultimately enhancing the quality and accessibility of mental health care.

References

Hugging Face. 2023. Training with lora. https://huggingface.co/docs/diffusers/en/training/lora. Accessed: 2025-04-15.

MEMO. Memo dataset. https://drive.google.com/file/d/1H2upxATpNS7x2mlNZSmybgJjtAhTTWRR/view. Accessed: 2025-04-15.

Nathan Raw. 2020. bert-base-uncased-emotion. https://huggingface.co/nateraw/bert-base-uncased-emotion. Accessed: 2025-04-15.

Aseem Srivastava, Tharun Suresh, Sarah P. Lord, Md Shad Akhtar, and Tanmoy Chakraborty. 2022. Counseling summarization using mental health knowledge guided utterance filtering. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, KDD '22, page 3920–3930, New York, NY, USA. Association for Computing Machinery.