

Summarizing Therapy Conversations

Introduction

This report explains how an automated tool was developed to summarize therapy conversations. The project focused on turning long therapy sessions into short, clear summaries, making it easier for mental health professionals to quickly see the main points. Using natural language processing (NLP) models, the tool picks out key ideas and emotions from each conversation. The report covers how the tool was set up, how it processes the data, the model used for summarization, and the results of this approach.

Methodology

The tool development involved multiple stages, including data acquisition, NLP model selection, and iterative testing. The workflow was built with the following components:

- **Dataset:** The dataset used in this project was accessed from https://huggingface.co/datasets/Amod/mental_health_counseling_conversations
- **NLP Model:** The summarization pipeline utilizes the BART model, a robust NLP model pre-trained for abstractive summarization tasks. This model is well-suited for identifying main ideas and summarizing textual data while preserving key information.
- **Data Preprocessing and Pipeline Setup:** The conversations consist of two main parts: *context* and *response*. These components are combined into a single input text, processed through the summarization model to generate a concise summary. The tool is implemented using Python with the transformers and datasets libraries.

Tool Implementation

To run the summarization tool, the following steps were followed:

- **Environment Setup:** A virtual environment was created to maintain a stable, isolated workspace for running the code. Necessary libraries, including transformers, torch, and datasets, were installed.
- **Summarization Pipeline Setup:** The BART model for summarization was implemented using the pipeline function from the transformers library. The script processed each conversation sequentially, summarizing and storing results in a text file.
- **Execution:** The script reads through each entry in the dataset, merges the context and response, generates a summary, and writes it to an output file, with real-time progress indicators for tracking.

Code Implementation:

```
import os

os.environ["HF_HUB_DISABLE_SYMLINKS_WARNING"] = "1"

from transformers import pipeline
from datasets import load_dataset

# Load the dataset
ds = load_dataset("Amod/mental_health_counseling_conversations")

# Set up the summarization pipeline
summarizer = pipeline("summarization", model="facebook/bart-large-cnn")

# Summarize and save each conversation
with open("summaries.txt", "w") as file:
    for i, entry in enumerate(ds['train']):
        text = f"{entry['Context']} {entry['Response']}"
        summary=summarizer(text,max_length=100,min_length=20,do_sample=False)[0]['summary.txt']
        file.write(f"Conversation {i + 1} Summary:\n{summary}\n\n")
    print(f"Summarized Conversation {i + 1}/{len(ds['train'])}")
```

Setup:

1) Create a Virtual Environment:

Bash:

```
python -m venv myenv
```

2) Activate the Environment:

- On Windows:

```
myenv\Scripts\activate
```

- On Mac/Linux:

```
myenv/bin/activate
```

3) Install Required Packages:

```
pip install transformers torch datasets
```

4) Run the Script:

```
python summarizer1.py
```

Results:

The tool successfully generated concise summaries for each conversation in the dataset, significantly reducing the volume of text while retaining essential information. The generated summaries enable mental health professionals to quickly assess session highlights, such as critical emotions, recurring themes, and notable changes in the client's tone or progress.

Example of Summarized Conversation Output

Original Conversation:

- **Context:**
The patient expresses feelings of worthlessness and struggles with self-esteem, often thinking about their perceived lack of value. They have not contemplated suicide but wish to change their feelings of worthlessness and are unsure how to do so.
- **Response:**
The response emphasizes that self-esteem is strongly influenced by social context. It suggests that the patient may benefit from surrounding themselves with people who uplift them rather than diminish their self-worth. It also advises the patient to seek inspiration from social media messages that reinforce a sense of purpose and reminds them that bad feelings are a natural part of life. The response suggests that feelings of worthlessness can motivate a person to seek positive changes in their life.

Generated Summary:

- **Summary:**

The social context in which a person lives greatly influences self-esteem. If the patient feels that everyone considers them worthless, they might need to find more supportive social connections. The response also encourages finding inspirational messages on social media to help reinforce a sense of purpose.

Link for summaries text file for

Example of Summarized Conversation Output

The summarized conversations generated by the tool are saved in a text file for further analysis and reference. You can view the full list of conversation summaries by accessing the file here:

<https://drive.google.com/file/d/1xfMo0-3aiOCubla7gIZgEN3tWyBqyU4j/view?usp=sharing>

Conclusion

This automated summarization tool offers a promising resource for mental health professionals, providing concise overviews of therapy conversations. By applying NLP techniques, the tool distills large volumes of text into short, manageable summaries that enhance therapists' ability to track client progress efficiently. Further development will focus on refining model performance and exploring additional NLP techniques to enrich the summaries.