

Dialogue summarization on BlendedSkill dataset

Task 1: Summarization

For nearly 1000-2000 conversations in BlendedSkill dataset, create ground truths to perform summarization and ensure the jist of the conversation is captured in the summary.

Task 2: Loss Calculation

1. Perform the summarization using any SOTA method and analyse how much information is lost in the summarization process.
2. Figure out a way to minimize this loss.

Task 3: Response generation

Make a chatbot (a working model on the terminal works) that summarizes (using the above loss-minimized summarizer) the previous history to give meaningful responses when the user interacts. So when the user enters something, the history is summarized and given to a model (choose from GPT-2/3.5/4, DialoGPT, T5, GPT-J, GPT-Neo) so it generates a response.

