

To, IITD-AIA Foundation of Smart Manufacturing

Subject: Weekly Progress Report for Week 6

Dear sir, Following is the required progress report to the best of my knowledge considering relevant topics to be covered.

What's happening this week:

- EDA-Pre-Processing the data i.e., scrapped out from site.
- Tokenization and stemming of messages/questions.

My understanding of INTP23-ML-01: CHATBOT for FSM

Scope: In the Chatbot for Site, I will try to scrape out all important info from website and according to the questions of users, it will provide best possible and accurate answers and it should follow the upcoming check points analyze user queries and understand user messages, provide answers to the messages of the user accurately, provide all information about the updated activities and save their time from searching on website and google search the information, provide answers accurately as much as possible, i.e., instead of providing irrelevant information, its' better to admit that bot is not able to understand the question.

Solution: In this, concept of deep learning either Pytorch or TensorFlow will be used along with some other libraries like Beautiful Soup, and one for scrapping the data on the website or using NLP and deep learning technique of transformer for question answering would increase efficiency of Chatbot. Using data of SQUAD, ROBUSTA model would be my preference.

Approach: Rule-based approach will be used for most common questions like hi, hello, bye. And web-scraping the data in order to give website related answers. For, questions/queries of users; we will use transformer (ROBUSTA model) and SQUAD dataset for pre-training and web-scraped data of IAFSM site for fine-tuning.

WEEKLY PROGRESS:

Day 37 (July 17- Monday)

there were some other errors like str does not have detach (), and after changing code there is error that batch size should match 512 dimensions. So, I will work on solving these errors. And today, my code is running. Initially, I directly finetuned the distil-bert on my data which is obviously very less to get much accuracy and It was printing almost whole context (provided) as answer. Then, I changed the model from distil-bert to distil-bert-squad i.e., a distil-bert model already finetuned with SQuAD dataset and also, I fine-tuned model on my data so accuracy is increased to great level. I made json file in Squad data format and fine-tuned the model, but while asking questions there are errors in that section.

Day 38 (July 18-Tuesday)

I was able to run the code and output was correct to great context. Today, I tried for using one complete context instead of providing chunks each time when we need to change the question (whose answer is not present in the earlier provided context). Also, after discussing with my mentor I got to know text2text conversion just searching for that for further understanding about text2text. **I made json file in Squad data format and fine-tuned the model, but while asking questions there are errors in that section.**

Day 39 (July 19- Wednesday)

While my model was accepting only 512 words token. Now, I have taken whole context of 23 k and divided into chunks of 512, now it's accepting whole context of 23k words. I splitted in length of 509 and added special token in the starting and end. Also, for question I added a separator after question marks sign. I made a whole context which is greater than 512 words actually it is 23k words

Day 40 (July 20- Thursday)

i did use the saved model and loaded the same in order to generate the answer. Also looked for gui as gradio for deloying my project

Day 41 (July 21 -Friday)

I'm working on for increasing the accuracy of my code. For that training the code again and again. Also, in order to deploy the model. I'm learning Gradio. How to provide input, output and how to deploy on hugging spaces etc. Also, I reduced the size of context removed repeated lines and unexplained terms and paragraphs from context in order to increase the accuracy. I tried different-different techniques like fine-tuning , training and reducing context in order to increase the accuracy.

Day 42 (July 22- Saturday)

I'm working on for increasing the accuracy of my code. For that training the code again and again. Also, in order to deploy the model. I'm learning Gradio. How to provide input, output and how to deploy on hugging spaces etc. Also, I reduced the size of context removed repeated lines

and unexplained terms and paragraphs from context in order to increase the accuracy. I tried different-different techniques like fine-tuning , training and reducing context in order to increase the accuracy. I used stride method i.e. stride of 50, max- length of 512 and then I'm asking questions.

Day 43 (July 23 -Sunday)

I'm working on for increasing the accuracy of my code. For that training the code again and again. Also, in order to deploy the model. I'm learning Gradio. How to provide input, output and how to deploy on hugging spaces etc. Also, I reduced the size of context removed repeated lines and unexplained terms and paragraphs from context in order to increase the accuracy. I tried different-different techniques like fine-tuning , training and reducing context in order to increase the accuracy. I used stride method i.e. stride of 50, max- length of 512 and then I'm asking questions. I got some problem in stride function instead from starting from last 50 words words from end it was taking next 50 words from previous start corrected that.