Department of Computer Science, University of Leicester CO7201 Individual Project

JokeBot: An AI Comedian Twitter Bot

Aakash Chahal
ac879@student.le.ac.uk

Interim Report

Project Supervisor: Dr Paula Severi
Second Marker: Muhammad Kazim

Word Count: 615 words
Date of Submission: 31/03/2023

1. Overview.

The has been significant progress made on the project Joke Bot, which came with both a lot of learning and some difficulties along the way, but the development of the project seems to be on the right track as all challenges and limitations have been pointed out and captured in the first half of the development of the project and these identified challenges hopefully would help build a better working final project.

2. Completed Tasks.

Following tasks are already completed:

i. Research: Research has been done on similar projects, APIs, libraries, and different techniques that can be helpful to build such a project.
Timeline: Weeks 1 and 2 were used to do the research for the project.

ii. Data Gathering and cleaning: Data (dataset of jokes) is already collected and has been cleaned since, the dataset consists of jokes cleaning of the dataset wasn't a difficult part it was mainly important to remove any unwanted characters or spaces.
Timeline: Weeks 3 and 4 were used to gather data and clean it.

iii. Language Models trained: Several models are trained on the jokes dataset(s), due to various limitations and challenges different language models and different libraries are used and being used to get a better result with available resources.
Timeline: Week 5 onwards
Models have been trained are producing output (joke) but to get a better output (joke) still testing out a few libraries / language models.

iv. Research on sentiment analysis: This is mainly a research related task which would be helpful when we have a final model to be tested. Since, the language model would be generating jokes the best way (other than testing) is to perform sentiment analysis and a few libraries like Pattern and Stanza upon research seems like a good option to perform sentiment analysis and get a good understanding of the generated jokes
Timeline: Week 6 (ongoing)
This task can't really be counted as completed until we finally test and analyse the outputs generated by the language model.

3. Tasks to be done.

Tasks that are next and yet to be completed are following:

i. Performing sentiment analysis: This task should and would be done right after the language model is trained as it's important to perform some sort of testing sentiment analysis in this case to know and measure that jokes generated by the model are actually funny and not showcase negative sentiments somehow, and the research on sentiment analysis and appropriate libraries is done. And the model if not performing well should be retrained with more dataset.

ii. Connecting to Twitter API: The next step after testing the language model and making sure it's generating desirable output (new jokes) is to build a program that connects to the twitter API which would allow us to post tweets and perform different functions like liking, retweeting, and commenting on a twitter post.

iii. Build an Interface between twitter API and the language model: This last task is the most important task as it is the main goal of the project to be able to tweet a newly generated joke on twitter. This is also the step where the recommended features like tweeting (a joke) around some trending topics and hashtags would be implemented.

iv. User feedback: User feedback would be gathered after the bot has been successfully deployed and is able to tweet and perform other activities on twitter, preferably if enough feedback gathered sentiment analysis can and should also be performed on the feedbacks to analyse the overall impact of the bot.

v. (Optional Task): Build a user interface (web app) that allows any user to generate a new joke and tweet it for them. This task is not included in the time plan because it is an optional task for now, and if there's enough time after the recommended tasks are done this feature would be implemented in the project.

4. Updated Time Plan

This is the updated time plan for the project considering the progress that has been made and the challenges that were faced:

Requirement	Start Date	End Date
Training and research - Training the model - Research on sentiment analysis	20/03/2023	02/04/2023
Performing sentiment analysis - Generate results (jokes) from the custom trained and pre trained model. - Perform sentiment analysis on both results. - Compare the results to decide what language model to choose	03/04/2023	16/04/2023
Connecting to twitter API - Making a new developer account. - Making a new profile for the bot. - Creating the private and public keys to access the twitter API.	17/04/2023	23/04/2023
Using twitter API - Use the API to automate the	24/04/2023	07/05/2023

tweeting, retweeting, and other activities on twitter. - Write a function to call the language model and get the result. - Work on other recommended features including likes, retweets, and comments.		
User feedback - Gather user feedback. - Perform sentiment analysis on the feedback.	08/05/2023	14/05/2023