DevOps Assignment Proposal

Student ID : ms20908188

Name : Thirunayan Dinesh Jeeva

1) Book Title

Building a continuously learning conversational AI pipeline

2) Introduction

The exponential progress in the development of deep learning-based language models has affected every possible natural language processing (NLP) and natural language understanding use-case to which conversational AI is no exception. Today's level –3 conversational AI development is far more complex and context driven than in the past and requires a lot of aspects to be considered apart from conversational paths like the conversational history, user's short-term conversational context and user's long-term conversational context.

Present conversational AI engines has 3 main important modules, the NLU (Natural Language Understanding) inference engine, the action executor and the response synthesizer. Within these 3 modules the NLU inference engine contains an ensemble of models to take in the user's raw message as input and classify any possible intents and entities and in addition also assigning a prior context score to influence and customize the next response. And chatbot development frameworks like Rasa enables developers to train these complex sets of models efficiently.

But on the other hand, one of the major bottlenecks in taking conversational AI solutions to production is the difficulty of continuously retraining the natural language models. Because here, there isn't just one model being retrained, it's a collection of interconnected models that must be trained parallelly. And continuous retraining is important to make that the chat assistant is more flexible and adaptable to even vague user questions. This continuous retraining pipeline should also be autonomous so that the chatbot can train itself without any explicit manual input from the engineers.

In this book, we are outlining a continuous retraining pipeline with CI/CD integrated, that would take in all the conversations a bot, including the mistakes the bot has made, preprocess the data and use the data to retrain itself. We

demonstrate how tools such as DVC (Data Version Control), Git/GitHub , Jenkins , AWS EC2 and AWS Lambda can be used to construct an autonomous continuous retraining pipeline with CI/CD integrated

3) Chapter outline plan

Chapter 1: Introduction to conversational AI

     a. Intro to building chatbots using Rasa

Chapter 2: Model Drift and why continuous learning is important

Chapter 3: Implementation

     a. Setting up rasa environment

     b. Training your first NLU chatbot
     c. Handling Fallback scenarios
     d. Adding DVC to separately manage datasets from code
     e. Deploying action-logic and inference models to production
     f. How to resolve the problem of data imbalance during retraining?
     g. Configuring AWS EC2 for continuous retraining
     h. Optimizing pipelines for performance and cost efficiency

Chapter 4 : Conclusion