

Question Answering Model Based Conversational Chatbot using BERT Model and Google Dialogflow

Nikita Kanodia

Intelligent Technology Innovation Lab
Victoria University
Melbourne, Australia
nikitakanodia25@gmail.com

Khandakar Ahmed

Intelligent Technology Innovation Lab
Victoria University
Melbourne, VIC Australia
khandakar.ahmed@vu.edu.au

Yuan Miao

Intelligent Technology Innovation Lab
Victoria University
Melbourne, VIC Australia
yuan.miao@vu.edu.au

Abstract— Contemporary conversational chatbots are user-friendly and possess the capabilities to simulate human conversations. However, they cannot evaluate large comprehensive datasets to provide an answer to the user. In contrast, the state-of-the-art Question Answering Model (QAM) trained on a large dataset can answer questions in the given context, and sometimes without context. This research designed a QAM to improve the customer's experience while using a chatbot for reading comprehension tasks using the BERT model and Google Dialogflow. QAM analyses and provides an accurate response using a comprehensive dataset and simulates human-like conversation. BERT model is used to predict accurate answers using the reading comprehension Conversational Question Answering (CoQA) dataset and Google Dialogflow to simulate human-like interactions. QAM extends the conventional way of using Google Dialogflow. A user-friendly Question Answering Model (QAM) reaps the benefits of Google Dialogflow and BERT Model integration. The BERT model and the chatbot interacts with each other using webhook and API. When a user interacts with the Dialogflow chatbot, it matches intents and sends the request to the BERT model. Finally, the BERT model provides an answer to the chatbot and respond to the end-user. The QAM provides accurate responses to end-users for questions based on large datasets.

Keywords— *BERT Model, Google Dialogflow, Question Answering Model (QAM), webhook, comprehensive dataset, chatbot*

I. INTRODUCTION

Chatbots are conversational applications used in many fields to help businesses interact effectively with their customers, such as a weather answering chatbot, food delivery chatbot, technical support chatbot, and university chatbot. These days chatbots are more user-friendly and helping businesses achieve better than ever customer service. There are many applications and technologies available to build chatbots. This research used state-of-the-art technology-Google Dialogflow to build a conversational chatbot. The chatbot simulates human conversation based on defined Intents and entities. However, conversational chatbots don't provide answers by analysing large datasets such as a Wikipedia passage, a document, or Conversational Question Answering (CoQA) dataset. With advancements in Natural Language Processing (NLP) models and technology, it has become possible to answer questions using large datasets. One such model used here is BERT Model to perform reading comprehension tasks. It outperforms the other NLP models by providing better accuracy in categorizing the identical words [1]. It takes the question and a passage into a single packed sequence and processes them based on the sum of token embedding and segment embedding to provide an accurate response.

Many researchers have developed NLP models question answering based on a reading comprehension passage. But there is a knowledge gap to extend the capabilities of NLP Question Answering to develop a user-friendly conversational chatbot. This research designed and implemented a model that can answer questions based on a large comprehensive dataset in a user-friendly and interactive way. For example, consider a scenario; some perspective university students have some questions about the university. With this proposed model, the university can input a passage or a document describing the university in the BERT model. Then, the QAM takes the user question, evaluates the passage to identify the answer, and responds to the student. In addition, the QAM understands the following conversational questions that might contain pronouns to refer to the words already used in the conversation. Thus, this research identifies QAM as an efficient model for reading comprehension tasks and providing the end-user with an accurate and interactive conversational question answering experience.

The QAM model implementation used a user-friendly chatbot in the front end to respond to the users' queries. The chatbot interacts with BERT Model using Fullfilment service in Google Dialogflow to respond to the customer and simulate general human interaction. The subsequent sections of this paper will provide the solution to design a conversational bot by integrating functionalities of the Chatbot with the advantages of the BERT Model. The literature review and critical analysis of the past papers identify the motivation for this research. This paper also discusses the implementation requirements and consists of a visual representation of the research method and the proposed user-friendly Question Answering Model (QAM).

II. RELATED WORK

In [2] implemented a chatbot using BERT Model and proposed a chatbot with more capabilities than a traditional chatbot to prevent misinfodemic and tackle the ever-growing concerns based on COVID 19. Their research used the text classification with BERT, which outperforms all other models. The user questions are categorized to achieve higher accuracy for the responses. Thus, forfeiting the challenge of BERT's maximum number of text strings as the model can now respond based on relevant question category and user question. The training and testing accuracy achieved is 98% and 96%, respectively. This research's results suggest that the BERT Model is the state-of-the-art model for question answering. It gives impressive accuracy and results when combined with text classification. However, the resulted model should be user-friendly to make this helpful chatbot. Moreover, it does not include any relevant integration to make it useful in the current pandemic situation and make the chatbot user-friendly. This research identified this gap and

implemented a model to overcome the challenge of existing user-hostile reading comprehension based question answering models.

III. QUESTION ANSWERING MODEL ARCHITECTURE

A. BERT Model

BERT is Bi-directional Encoder Representation from Transformer. BERT model is trained on large Wikipedia and book corpora, which saves a lot of heavy computation and time involved in training the model. This paper uses a fine-trained BERT Model from the Hugging face transformer for various NLP tasks such as Question Answering, Sentiment Analysis, and Named Entity Recognition. For the Question Answering, it uses a start and an end vector. It takes the question and a passage into a single packed sequence and processes them based on the sum of token embedding and segment embedding. It uses Transfer Learning, which is already pre-trained on a vast dataset and fine-tune for specific tasks. It also uses word piece tokenization to break down the words, enhancing the training performance and helps understand the word's context. For example, there are three words 'help', 'helping,' and 'helper'. The model divides the word into 'help' and relevant suffix, i.e. 'er' and 'ing'. Its embedding outperforms any other model as it understands the word's context better than other NLP models [3]. The model learns the meaning of the word help and includes the context of the rest of the word in the suffix, to use by the number of different words containing similar suffix [4].

Unlike other directional models, it uses a transformer encoder to read the entire word sequence in a go. Thus helping the model understand the word's accurate contexts considering the left and right of the word. The pre-trained model can be used to fine-tune for the desired task.

B. Google Dialogflow

Google Dialogflow is one of the most reputed and best chatbots available in the market. The chatbot answers simple interaction questions. And for a question based on a comprehensive dataset, the Dialogflow interacts with the BERT Model using fulfillment service, webhook and flask API to provide the answer. Below [5, Fig. 1] represents the flow of Google Dialogflow.

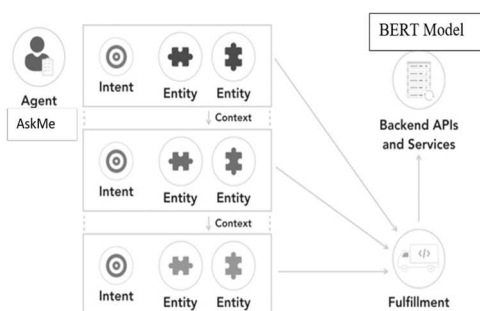


Fig. 1. Working of Google Dialogflow Chatbot [5]

C. Integration

A webhook is an HTTP POST request called upon meeting specific criteria. The QAM uses Python Flask API to create a webhook. In Dialogflow, webhook calls function in the specified server as indicated by a matched intent. Flask API help add the functionality to connect Google Dialogflow

chatbot with the BERT Model implemented in Jupyter Notebook local machine and send HTTP requests. The research also uses ngrok to create a tunnel between Jupyter Notebook local machine and Google Dialogflow on the internet. It helps flask API to connect with the Dialogflow intent. The Question Answering Model effectively integrated the BERT model, Google Dialogflow, python flask API, and webhook URL.

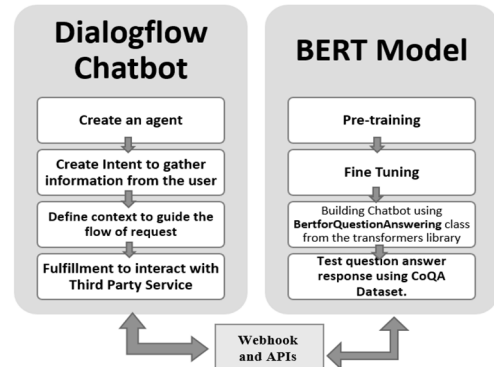


Fig. 2. Question Answering Model (QAM) Methodology

IV. QUESTION ANSWERING MODEL IMPLEMENTATION AND RESULT

The research used the Conversational Question Answering (CoQA) dataset to fine-tune the pre-trained BERT model. It is an extensive dataset for conversational reading comprehension. The benefit of using the CoQA dataset over any other dataset is that it includes both passages and interrelated questions that appear as a conversation. Thus, this dataset is suitable for this research as the research focuses on reading comprehension and conversation tasks. First, the dataset is loaded from the CoQA dataset (i.e., coqa-train-v1.0.json) to the BERT model [6]. Then, the dataset's JSON format is cleaned and loaded into a data frame for further processing.

For the Question Answering Model, the BERT Model is implemented using Hugging face transformers and necessary libraries such as numpy, pandas, torch, BertForQuestionAnswering, BertTokenizer and fine-tune using CoQA dataset. Then python flask API is added, followed by the creation of Dialogflow agent and intent. Then, training phrases are included (Fig. 4). These training phrases are the questions based on the text passage fed into the BERT model [7, Fig. 5]. The passage used in the BERT model contains information about Victoria University. Then ngrok is used to create a tunnel between the Jupyter Notebook where the BERT model is executed to Google Dialogflow on the internet. Finally, fulfillment service is used in the Dialogflow to integrate the Dialogflow with the BERT model using webhook URL and Python flask API.

When a user asks questions based on predefined training phrases and the passage included in the BERT Model, the Dialogflow agent calls a webhook request to the BERT Model

to get a webhook response. As a result, the user receives the answer to the question.

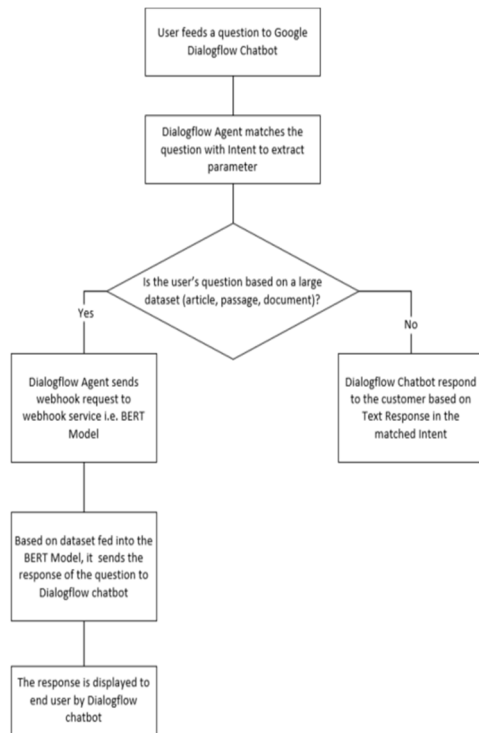


Fig. 3. Question Answering Model (QAM) Implementation

The test shows successful results for conversational questions asked based on the specific article. For example, consider the following dataset and set of training phrases:

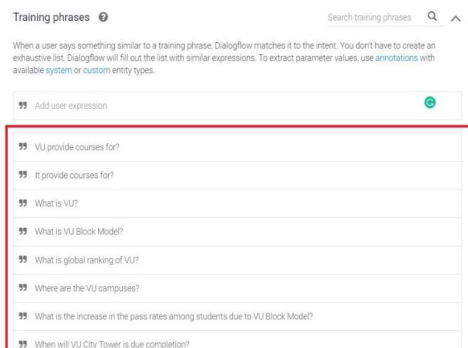


Fig. 4. Training Phrases based on the Passage fed into Bert Model

"Victoria University (VU or Vic Uni) is a public research university based in Melbourne, Victoria, Australia. It is one of only six dual-sector universities in Australia, providing courses in both higher education and Technical and Further Education (TAFE). 2016 marked VU's centenary as an educational institution and its 25th anniversary as a university. The university has several campuses in Melbourne Central Business District, Melbourne Western Region, and in Sydney, comprising six academic colleges, six research institutes, seven research centres and VU's Victoria Polytechnic (providing vocational education and training). The new VU city Tower will be Melbourne's tallest vertical campus and is due for completion in 2021. It also offers courses at partner institutions throughout Asia. Victoria University Melbourne is a globally recognised institution, ranking in the top 2% of universities worldwide (2021 Times Higher Education World University Rankings), and 56th globally (2nd in Victoria) in the 2020 Times Higher Education (THE) Young University Rankings. VU was also ranked 11th in the world (1st in Australia) for Peace, Justice and Strong Institutions (Times Higher Education Impact Rankings 2021). In 2018, Victoria University became the first university in Australia to adopt the block model style of teaching for all undergraduate courses. Under VU's Block Model, students' study and complete one unit at a time over a four-week period (a block), working collaboratively in smaller classes. Unlike the standard model of tertiary education in Australia, students studying under VU's Block Model focus on learning one unit every four weeks, rather than juggling multiple units and assessments at the same time within a semester. Since introducing the VU Block Model, pass rates amongst onshore undergraduate students have increased, with over 90% of students passing their enrolled units in 2020. Overall grades for this cohort has also grown, with over 60% receiving distinction levels or higher in the same year. This improvement in academic results is complemented by higher levels of learner engagement amongst its students. The latest Australian Government 2020 Student Experience Survey (SES) has Victoria University now ranked by students as the top University in Victoria, and third overall in Australia, for learner engagement.

Fig. 5. Passage fed into the BERT Model [7]

The QAM model recognizes the context of words used in the conversational language in this article based on the previous question, indicating a successful implementation of the question-answering BERT Model chatbot. Fig. 6 shows the questions input in the Dialogflow and their responses.

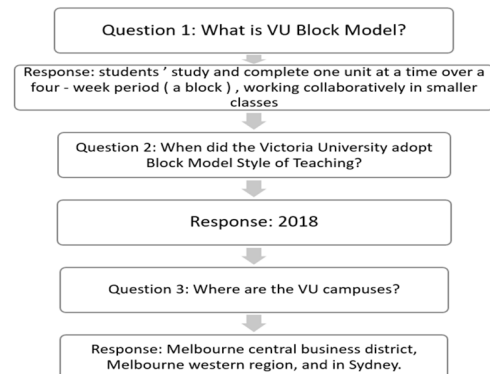


Fig. 6. Question- Answering Result

The QAM is successfully implemented, as shown in Fig. 7. The agent AskMe answers the questions based on the Wikipedia passage about Victoria University into the BERT model. The QAM implementation can be used for different use cases to meet specific business objectives.

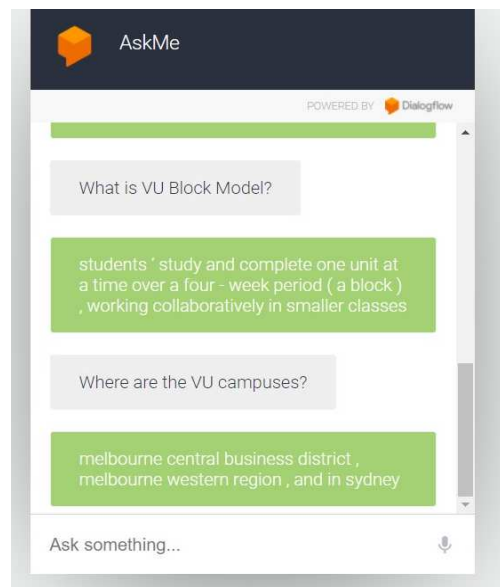


Fig. 7. Successful Implementation of Question Answering Model (QAM)

V. CONCLUSION

This research successfully implemented the Google Dialogflow intents for demo chat and BERT Model for reading comprehension tasks. Along with the complete setup of the environment and fine-tuning of the BERT Model using the CoQA Dataset. The research identified flask API, webhook, and ngrok as appropriate for integrating the BERT model and Google Dialogflow chatbot. The python flask API successfully integrated BERT Model with Google Dialogflow. The user can get answers from the Dialogflow based on the passage fed into the BERT model.

This paper also provides a basic understanding of the BERT Model and Google Dialogflow chatbot, CoQA dataset,

flask API, and webhook and justifies their selection. In addition, the knowledge gap in the current work is justified and provided a successful demo of the BERT model. Users can now deploy the Question Answering Model (QAM) in multiple disciplines to fulfil their objectives. Thus, the research demonstrated the extensive capabilities of the chatbot by extending its ability to answer questions based on comprehension and provide a user-friendly interface.

REFERENCES

- [1] S. Ozan and D. E. Tasar, "Auto-tagging of short conversational sentences using natural language processing methods," in *Proc. 9th Signal Processing and Communications Applications Conference (SIU)*, 2021, pp. 1-4. Accessed on: Aug. 14, 2021. [Online]. doi: 10.1109/SIU53274.2021.9477994.
- [2] E. Amer, A. Hazem, O. Farouk, A. Louca, Y. Mohamed and M. Ashraf, "A proposed chatbot framework for COVID-19," in *Proc. International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC)*, 2021, pp. 263-268. Accessed on: Aug. 10, 2021. [Online]. doi: 10.1109/MIUCC52538.2021.9447652.
- [3] H. Khorashadizadeh, R. Monsefi and A. Foolad, "Attention-based convolutional neural network for answers selection using BERT," in *Proc. 8th Iranian Joint Congress on Fuzzy and intelligent Systems (CFIS)*, 2020, pp. 121-126. Accessed on Sep 3, 2021. [Online]. doi: 10.1109/CFIS49607.2020.9238669.
- [4] C. Khanna, *Question Answering with a Fine-Tuned BERT*, Towards Data Science. Accessed on: Aug. 14, 2021.[Online]. Available: <https://towardsdatascience.com/question-answering-with-a-fine-tuned-bert-bc4dafd45626>
- [5] LinkedIn Learning, *Building Chatbots Using Google Dialogflow*, Dec. 11, 2019. Accessed on: March. 1, 2019. [Video file]. Available: <https://www.linkedin.com/learning/building-chatbots-using-google-dialogflow/breakdown-of-dialogflow-components?u=22952571>
- [6] S. Reddy, D. Chen and C. D. Manning, *Coqa: A conversational question answering challenge*, 7, MIT Press, 2018. Accessed on: Aug 10, 2021. [Online]. Available: <https://arxiv.org/abs/1808.07042>
- [7] "Victoria University, Melbourne," Aug. 31, 2021. Access on Sep. 18, 2021.[Online]. Available: en.wikipedia.org/wiki/Victoria_University_Melbourne