International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 07 | July 2021 www.irjet.net p-ISSN: 2395-0072

University AI Chatbot

Aman Mishra[1], Bahram Hashmi[2], Deepak Kumar[3], Himalaya Sagar[4]

[1-4]UG Student, Dept. Information Technology, MIT School of Engineering, MIT Art Design and Technology University, Pune, Maharashtra, India.

Abstract - *Chatbot, or conversational interfaces as they are also called, introduce a new way for individuals to interact with computer systems. Initially, to answer a question the software program uses a search engine, or filling out a form. A Chabot allows a user to simply ask question in the same way that they would deal with a human. However, the deployment of the Chatbots has been increased exponentially on online chat platforms. The technology at the center of the rise of the chatbot is natural language processing (NLP). The advancement in machine learning has immensely enhanced the accurateness and efficacy of natural language processing, making chatbots a feasible option for numerous organizations. But yet to achieve many tasks there is need to make chatbots as much efficient as possible. To address this problem, in this paper we provide Deep learning can be applied on intent classification algorithm to classify and find patterns in the natural language, using word embedding. We need to provide training set of a processed data with, and it will pick up patterns in data and classify the intent accurately and in fairly less amount of time*.

Key Words: **Natural Language Processing (NLP), Deep learning, Chatbot, Intent Classification, Artificial Intelligence (AI), Natural Language Tool Kit (NLTK), Convolutional Neural Network (CNN).**

1. INTRODUCTION

Chatbots are not a recent development. They are an artificially created platform which can understand human language, process it besides interact back with humans while performing specific tasks. For example, a Chatbot can be employed as a helpdesk executive. The very first Chat system or chatbot was created by Joseph Wiesenbaum in 1966, named Eliza. These all materialized when Alan Turing published an article named "Computer Machinery and Intelligence", and raised an intriguing question," Can machine think?", and from then, we have perceived improved numerous chatbots came into existing which are more naturally conversant and technologically advanced. These progressions have led us to an epoch where tête-à-têtes or discussions with chatbots have become as normal and natural as with another human.

Today, almost all companies have Chatbot to engage their users and serve customers by catering to their queries. As per a report by Gartner, Chatbot will be handling 85% of the customer service interactions by the year 2020. Also, according to the advancements in technology experts have predicted that more than 80% of businesses are likely to have some type of chat system or Chatbot automation by 2020 (Outgrow, 2018). We virtually will have chatbots all over the platforms, but this doesn't inevitably mean that all will be well-functioning. The challenge here is not to develop a chatbot, but to develop a Chatbot with capability to classify and give accurate responses for the Queries.

NLP is an applied artificial intelligence (AI) program that helps your chatbot to analyze and understand the natural human language communicated by your customers.NLP helps chatbots understand, analyze and prioritize the questions according to the density & this enables bots to respond to customer queries faster than a human being. With the help natural language processing, we are apt to "train" our chatbot with the various interactions it will go through, and help streamline the responses. A deep learning Chatbot learns precisely from scrape through a procedure called "Deep Learning." In this process, the chatbot is created using machine learning algorithms. A Deep learning chatbot acquires everything from its data and human-to-human dialogue. In this paper we provide Deep learning can be applied on intent classification algorithm to classify and find patterns in the natural language, using word embedding.

2. Materials and Methods

2.1 Intents Dataset

First, we define our Intent Dataset based on tags, Patterns and Responses for each of them. The Intent Dataset can be created in a format from where our Chabot will classify the Tag which the User will ask and then will check for corresponding Pattern associated with it and then will give the corresponding Response associated with it.

- Tags: is the single word string that our Chatbot will classify according to the Query asked by the User in the Input Field. In that Query our Chatbot will try to find the related Tag associated with that Query in our Dataset and will match that Query to that particular Tag.
- Patterns: are the multiple similar sentences that are present in that particular Tag to match with the Query and associate the most suitable pattern for that Query. The Chatbot will search and relate the Query with one of the best matching Pattern.

Responses: are the suitable answers to the related query that was asked by the user. After matching with associated Tag and Pattern of Query our Chatbot will fetch the Response from Dataset and give the suitable Response back to the User for that Query.

2.2 Methodologies for Problem Solving

The Chatbot takes the input from the User and passes this input to a Open Source Library called as GingerIt .The GingerIt Library has a method for Auto-Correction of the Input provided by the user if the Input has spelling or Grammatical Mistakes. The Input Query is passes on to the Natural Language Tool Kit (NLTK) which works on Intent classification Algorithm to identify patterns in the Query using word embedding techniques and classify the Tag associated with the Query. The Query is then passed on to the CNN Model which is been trained on a processed data and it will identify the patterns in data and classify the intent precisely and in impartially a lesser amount of amount of time. The CNN model will then predict the suitable response associated with the Query and replies back to the User.

- GingerIt: It is an Open-source Library which is basically a Natural Language Processing which wrapper around the gingersoftware.com API and it is used in the Chatbot for Auto-Correction of the Input provided by the user if the Input has spelling or Grammatical Mistakes.
- Natural Language Processing (NLP): is a component of Artificial Intelligence (AI) which has ability to understand human language which referred to as Natural Language. In this project we have used various Natural Language Processing technique to identify human Language so that our CNN model easily predicts the response of the Input Query.

  The techniques which we have used are:

  i. Tokenization: It is an NLP technique which helps to break the raw text in words called tokens, we can understand it by an example: Suppose we have a sentence called ['Natural Language Processing'], after tokenization we will get the tokens as ['Natural', 'Language', 'Processing'], these token helps to understand the context led to the development of NLP model.

  ii. Stemming and Lemmatization: These are the NLP techniques which help to analyse the meaning of the word and the context in which the words are said. Stemming creates the root form of the word. We can understand through an example we have a set of word ['Studying', 'Study', 'Studies'] stemming breaks down these words into a root word called ['Studi'] due to which a unique word has been formed.

  iii. Bag of words: It is an NLP technique which is done with the help of NLTK called count vectorizer in which the each word is represented as the real-valued vector in a high-dimensional space. The vectors are represented in such a way that have similar meanings will have similar representations in the vector space. These Vectors get updated during the model Training.

- Convolutional Neural Network (CNN): The CNN model is a class of Deep Neural Networks and we have used the CNN model with 3 Dense fully-connected layers. First layer with 128 neurons and input of size of number of patterns with ReLU as activation function. Second layer with 64 neurons and activation function ReLU. 3rd output layer contains number of neurons equal to number of intents to predict output intent with softmax as Activation function. These parameters are a result of several iterative processes, with the aim being to get the best model architecture for the dataset at hand. The System Flow Diagram and CNN Architecture will be as follows in Fig.1 and Fig.2 respectively.

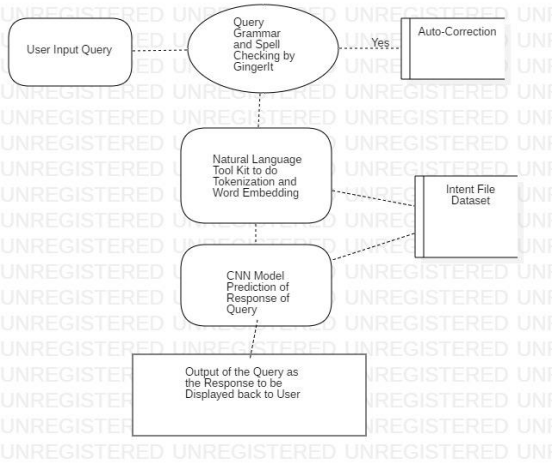

Fig. 1: System Flow Diagram

```
model created
Model: "sequential_2"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
dense_5 (Dense)              (None, 128)               10624
_________________________________________________________________
dropout_4 (Dropout)          (None, 128)               0
_________________________________________________________________
dense_6 (Dense)              (None, 64)                8256
_________________________________________________________________
dropout_5 (Dropout)          (None, 64)                0
_________________________________________________________________
dense_7 (Dense)              (None, 32)                2080
_________________________________________________________________
dropout_6 (Dropout)          (None, 32)                0
_________________________________________________________________
dense_8 (Dense)              (None, 20)                660
=================================================================
Total params: 21,620
Trainable params: 21,620
Non-trainable params: 0
_________________________________________________________________
```

Fig. 2: CNN Model Architecture

3. RESULTS AND DISCUSSION

The User is asked to insert a Query into the text box which is then forwarded to GingerIt Library for Auto -Correction in spellings or grammatical mistakes in the Query, if it exists then it's Auto-corrected or else the same sentence is passed on to the Natural Language Tool Kit(NLTK) for the pattern recognition and word embedding for using as a pre-processed data for our CNN model to predict the Output

Response from the CNN Model that is trained on our Dataset and gives back the Response to the User in the Text Box.

The CNN model learns from these patterns and gives the best suitable Response to the user according to the pattern of asking Queries by the User. We have some results for the Auto-Correction of the misspelled Query and its corrected Output as follows with the pattern of Responses generated by our CNN based Chatbot are as follows.

Screenshots of the Auto Correction of Misspelled Query and proposed system of Chabot are given in Fig.3 and Fig.4 respectively.

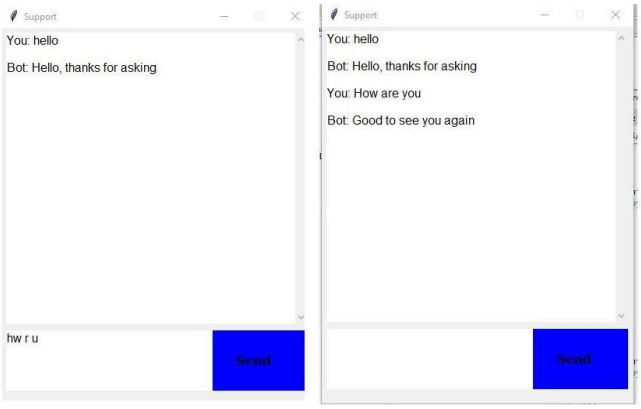

Fig.3: Auto-Correction of Misspelled Query.

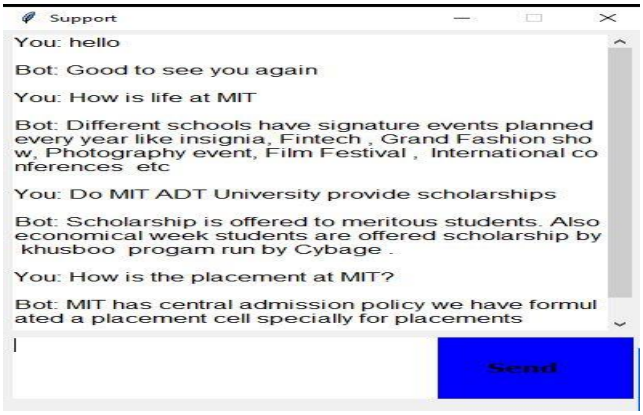

Fig.4: Queries and Responses from Chatbot.

4. CONCLUSIONS

AI Chatbots are drastically changing businesses needs and fulfilling the requirements within a fraction of second by responding to the user's query and providing them solutions. There is a widespread platform of chatbot development that is available for various inventiveness, such as e-commerce, retail, banking, leisure, travel, healthcare, and so on.

Chatbots can influence out to great onlookers on messaging apps and be more operative than humans. They may advance into accomplished information-gathering means in the near future. We have developed an AI chatbot using CNN model and Natural language processing for our university. If a new user wants acquire about our university whether they wanted to know about infrastructure or campus placement or any admission related queries they can simply chat to our user friendly chatbot which replies to the user within a second with the best possible answer. The user may get satisfied with our chatbot reply and get to know more about our university. The chatbot is efficient and timesaving for the user who wants to enquire about the university.

REFERENCES

[1] Challamalla Balaji Ram Mohan, Ankush Babu Divi, Abbineni Venkatesh, Bandla Sai Teja, Mohan Krishna Kotha "Chatbot for University Resource Booking ", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN: 2456-3307, Volume 5 Issue 2, pp. 113-116, March-April 2019.

[2] B. R. Ranoliya, N. Raghuwanshi and S. Singh, "Chatbot for university related FAQs," 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), IEEE 2017, pp.1525-1530, doi: 10.1109/ICACCI.2017.8126057.

[3] Yogi Wisesa Chandra, Suyanto Suyanto,"Indonesian Chatbot of University Admission Using a Question Answering System Based on Sequence-to-Sequence Model",Procedia Computer Science, Volume 157,2019,Pages 367-374,ISSN 1877-0509.