IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-7, Issue-10, Oct 2019 E-ISSN: 2347-2693

Virtual Assistance Using Chatbots

Kavitha K V[1*], Athira S Pillai[2], Disha Dinesh Majithia[3]

[1,2,3] Dept. of Computer Science and Engineering, Sree Chitra Thirunal College of Engineering, Thiruvananthapuram (Kerala)-695018, India

[*]*Corresponding Author: kavitha279@yahoo.co.in*

DOI: https://doi.org/10.26438/ijcse/v7i10.7780 | Available online at: www.ijcseonline.org

Accepted: 12/Oct/2019, Published: 31/Oct/2019

***Abstract*—** In the recent era of technology, chatbots is the next big thing in the epoch of conversational services. It acts as the ultimate virtual assistant for completing various tasks ranging from answering queries, getting driving directions, turning up the thermostat in smart home, to playing one's favorite tunes etc. Whenever faced with unchartered territories humans have a tendency to get anxious and it is but natural- and even more so in the case of students who are about to join a new course. They are always filled with endless queries and mostly all of them are similar across people. Rather than flooding a single person with all those and to answer them over and over again, it is apt to segregate usual common queries and their solutions at a single place which is easily accessible by all. In this paper an efficient Chabot is designed based on neural network, which provides an accurate answer for any query based on the dataset of FAQs. This Chabot can be used by Sree Chitra Thirunal College of Engineering to answer frequently asked questions to curious students in an interactive fashion. The accuracy of the bot was noted to be 0.78 with a loss of 2.06.

***Keywords*—**Artificial Intelligence, NLP, ML

I. INTRODUCTION

Chatbots is a virtual person who can successfully speak to any human being using interactive textual skills. Currently, many cloud based chatbot services are available for the development and enhancement of the chatbot sector such as IBM Watson, Microsoft bot, AWS Lambda, Heroku and many others. A simple chatbot is not a challenging task as compared to complex chatbots and developers should understand and consider issues such as constancy, scalability and flexibility issues along with high level of intention on human language. Recent advancements in machine learning techniques may able to handle complex issues related to chatbots.

Chatbots can be defined as software agents that converse through a chat interface. Now, what that means is that they're software programs that are able to have a conversation, which provides some kind of value to the end user. The user can interact with the chatbot by typing in their end of the conversation, or simply by using their voice, depending on the type of chatbot provided. Virtual assistants like Apple Siri or Amazon Alexa are two examples of popular chatbots interacting via voice rather than text. Typically, the chatbot will greet the user and then invite them to ask some kind of question. When the user replies, the chatbot will parse the input and figure out what's the intention of the user's question. Finally, it will respond in some form of consequential or logical manner, either providing information or asking for further details before ultimately answering the question. Great chatbots can keep up this back and forth in a natural way, within the scope of what the chatbot is designed to do. They make the user feel understood and helped. They create a certain rapport with the user, without pretending to be human. Very often chatbots offer some form of virtual assistance to the user. The first historic chatbots in the late sixties and early seventies were in fact mostly an exercise in coming up with seemingly intelligent and human-like chatbots with relatively simple technology.

A dominant method to investigate the patterns and future trends of a topic related to research is bibliometric analysis. This method is used by many researchers to explore the research trends in different areas, such as the role of IT innovation, project management, research and business intelligence and analytics. Three tools used for bibliometric analysis are: (i) Literature databases - Web of Science (WoS) and ProQuest: to collect publication information and citation report, (ii) CiteSpace: to evaluate and cluster data, and (3) Bibliometrix: to find out co-occurrence patterns

II. OBJECTIVE OF CHATBOT DESIGN

Presently, SCTCE has no provision for any such system and therefore the suggestion to create one to simplify the

redundant nature of work and utilize manpower efficiently. The need for a virtual assistant is imminent in this scenario. The bot aims at solving all admission related queries of the students. It is possible that it is not just students that are trying to access the website chatbot; there is also a possibility that people who want to know about the staff qualification at the college access it. Solving such queries is also what the bot aims at doing efficiently.

III. RELATED WORK

"Chatbot and conversational agents: A bibliometricanalysis"[1] tells about the increasing demand of chatbots. "They are replacing some of the jobs that are traditionally performed by human workers, such as online customer service agents and educators. Due to rapid development in the field of artificial intelligence (AI), chatbots performance keeps improving. Chatbots can nowadays "chat" like a human being and they can learn from experience." it quotes."An e-business chatbot using AIML and LSA"[2] explores the possibilities of a chatbot being used in the business world. The paper has stated the difficulties of the accuracy with which an online chatbot would converse by copy pasting pre written answers. It usesAIML and LSA to create the chatbot unlike our chatbot."Programming challenges of chatbot: Current and future prospective" [3] speaks about all the challenges that are encountered while building a chatbot. Chatbot building is a tedious task. It speaks of the scalability, stability and flexibility issues."Chatbot using knowledge in database human-to-machine conversation modeling"[4] gives the insight to chatbot building. It explains the sentence similarity algorithm and concepts of pattern matching."Chatbot for university related FAQs"[5] is the paper that has done work similar tothe present project. ManipalUniversity has come up with a chatbot for their university. The bot is a FAQ bot and only college related FAQs can be asked. They have used AIMLto build the bot."Evaluation of Modern Tools for an OMSCS Advisor Chatbot"[6] At Georgia State, "Pounce" the chatbot has been helping students navigate the application process. At technical University of Berlin, Germany, a chatbot named "Alex" helped students in test groups and schedule classes more efficiently.

IV. METHODOLOGY

In this work the development of an interactive chatbot for College related queries have been done. The steps in the design of chatbot are as follows.

- The query is posted by the user on the chatbot.
- Based on query of the user processing is done to match the predefined format by the developer.
- Perform a pattern matching between user entered query and knowledge (pattern).

Finally the user is presented with pattern based answer to answer their queries. This chatbot is designed for educational sector, where a student or a parent can ask query regarding college related matters like admission, college location, entrance rank and other things related to academics. The query is posted by the user on the chatbot and response is generated based on pattern matching techniques.

A. *Chat Interface*

Chat interface is the medium used by the chatbot to connect with the user. The chatbot can be embedded in the official website of the college which is www.sctce.ac.in. HTML and CSS was used in the interface.

B. *Architecture*

The architecture model of chatbot is decided based on the core purpose of development. Chatbot has two types of responses: it can either generate a response from scratch using machine learning models or use some heuristic to select an appropriate response from a library of predefined responses [7]. The problem with pattern-based heuristics is that all the patterns need to be programmed manually. This is a tedious task, since the chatbot has to distinguish hundreds of intents for different scenarios. For training the bots, Intent classification is completely based on machine learning technology. With a training set of thousands of examples the chatbot can be trained to pick up patterns of data and learn from it. Scikit-learn are a popular machine learning library that helps in executing machine learning algorithms.

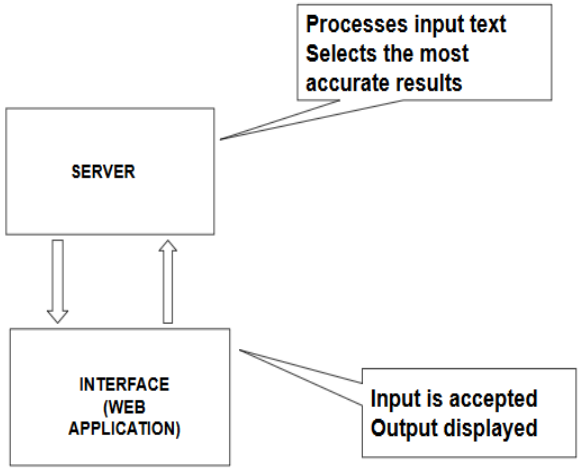

Fig 1: Chatbot Design

The chatbot architecture shown in the Figure1 and Figure 2 describes how the various components of the chatbot can be integrated together to form the final system. It consists of a channel which is used to interact with the user. The channel could be any software that will run on a system or if the chatbot is implemented as a web application, the website will act as the channel. The channel will be connected to the system backend which is the main part of the entire system that will be used to process the queries that the user sends to

the system. This query is passed on to the data processing block of the system which will implement the algorithm used in the system to get the desired results. The keyword matching algorithm used here will attempt to identify keywords in a sentence. If one or more keywords are found in the user's input text then an answer will be retrieved. The database or the dataset used here contains the required information that will be parsed by the chatbot to obtain the desired results.

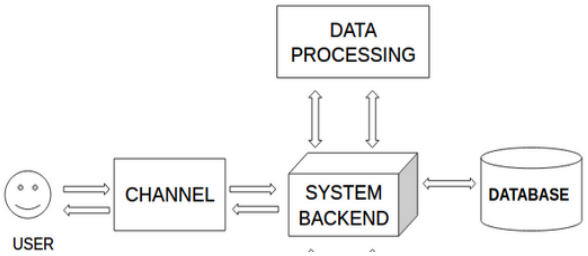

Fig 2: Architecture of proposed system

Dataset was made with the help of college students. Google forms were distributed and survey was carried out.

C. Algorithm

Semantic similarity is calculating score based on semantic relation between two sentences or strings. Given two sentences or strings, similarity can be determined by measuring it. The higher the score value the more similar is the meaning of two sentences. The score lie from 0 to 1.The algorithm is provided below.

Sentence similarity measurement algorithm:

Step 1: Start

Step 2: Divide each sentence into a set of tokens

Step 3: Implement bigram on the divided sentence

Step 4: Apply the following equation to the tokens:

$$\frac{\partial(stm1 \in stm2) \cup \partial(stm2 \in stm1)}{\partial(stm1) \cup \partial(stm2)}$$

Step 5: Obtain the similarity score for both the sentences.

Example: Let's consider two words "burp" and "burrup" then

stm1= "burp", stm2= "burrup"

After dividing both the sentences into bigrams we get:

stm1= 'bu','ur','rp' = 3

stm2= 'bu','ur','rr','ru','up' = 5

stm1s2 = 'bu','ur','rp' = 3

stm2s1= 'bu', 'ur' = 2

The sentence similarity score:

$$\frac{3 \cup 2}{3 \cup 5} = \frac{5}{8} = 0.625$$

V. RESULTS AND DISCUSSION

The main idea behind chatbot implementation is to fulfill the academic needs of the users. . But there are two competing concerns: if the training data is less, the parameter estimates have greater variance. If testing data is less, the performance statistic will have greater variance. The data should be divided such that neither variance is too high, which is more to do with the absolute number of instances in each category rather than the percentage. The neural network that was developed for the model consists of various metrics that are to be used to estimate the accuracy and loss of the model. The various factors that affect are number of neurons in the model, the number of epochs used to train the model, the batch size of input provided and the error threshold value that is set. Here, epoch defines the number of times the training data has to be passed through the network, batch size defines the number of inputs that is processed at a time by the neural network and error threshold defines the probability value below which model prediction error increases.

The model was trained using 8 neurons in hidden layer and 16 neurons in the hidden layer separately and the variations in the values of loss and accuracy were checked.

Table 1: Loss and accuracy for error threshold 0.02

No. of Epoch	8 neuron layer		16 neuron layer	
	Loss	Accuracy	Loss	Accuracy
1000	2.51	0.650	**2.06**	**0.78**
2000	2.58	0.651	3.52	0.50
4000	3.25	0.517	2.49	0.70

Table 2: Loss and accuracy for error threshold 0.01

No. of Epoch	8 neuron layer		16 neuron layer	
	Loss	Accuracy	Loss	Accuracy
1000	3.06	0.56	**2.7**	**0.71**
2000	2.81	0.57	3.04	0.51
4000	3.11	0.51	2.5	0.68

It was observed that on changing the epoch values for 8 neuron layer network, the loss increases and accuracy decreases as shown in Table 1. This proved that an eight neuron network is not suitable for carrying out the model training. Thus, we carry out the same tests on the sixteen neuron layer network to get the outputs as shown in Table 1. It was observed that when the number of epochs was made 1000 the loss obtained decreased with a considerable increase in accuracy. Table 2 gives the values of loss and accuracy when the error threshold considered was 0.01 and the considered batch size of inputs was 200. In the case of 8 neuron network it is observed that accuracy obtained is less. As per Table 2, it is seen that the model under fits data for 2000 epoch values and over fits data for 4000 epoch value. After estimating accuracy and loss for various values the best metric combination was chosen as epoch value 1000 for the amount of dataset the model is to be trained on.

Amongst the total instances, an 80-20 split validation was carried out and it was seen that for every test set of queries provided each time as input there were correct responses provided 70% of the times. While the deep learning neural network was trained, it was seen that during each training there was a loss of 2.06 and an accuracy of 0.78.

VI. CONCLUSION

Chatbots needs to be incorporated more into websites of colleges so that information can be passed on to the users without them having to go though the entire websites. This chatbot helps the students of SCT College of Engineering to get information regarding college without going through the website content and thus making human work less. It reduces tedious work that usually a user would have to carry out if they need to get any information from the website of the college. As there is no manpower employed in case of chatbot usage, it makes it easy for the users to get the required results of the queries. The accuracy of the chatbot designed using neural network was 78%. This chatbot can be made more efficient by incorporating machine learning algorithms in the future. .

REFERENCES

[1] C.B.Lee, H. N, "*Chatbots and Conversational Agents: A Bibliometric Analysis*", IEEE International Conference on Industrial Engineering and Engineering Management, **2017**.

[2] Thomas N T Amrita, " *An E-business Chatbot using AIML and LSA*", Intl. Conference on Advances in Computing, Communications and Informatics, **2016.**

[3] A M Rahman, Abdullah Al Mamun, Alma Islam, "*Programming challenges of Chatbot: Current and Future Prospective*" ,IEEE International Islamic University Chittagong, **2016**.

[4] BayuSetiaji, Ferry WahyuWibowo , "*Chatbot Using A Knowledge in Database Human-to-Machine Conversation Modeling*" International Conference on Intelligent Systems, Modelling and Simulation, **2016**.

[5] Bhavika R. RanoliyaNidhiRaghuwanshi, Sanjay Singh Department of Information and Communication Technology Manipal University, "*Chatbot for University Related FAQs*" , International Conference on Advances in Computing, Communications and Informatics, **2017.**

[6] Eric Grigori, "*Evaluation of Modern Tools for an OMSCS Advisor Chatbot*" Georgia Institute of Technology Atlanta, **2017**.

[7] Luis Fernando D'Haro and Rafael E. Banchs, "*Learning to predict the adequacy of answers in chat oriented human agent dialogs"* IEEE Region 10 Conference (TENCON), **Malaysia, 2017.**

Authors Profile

Kavitha K V pursed Bachelor of Technology and Master of Technology from Kerala University, India. She is is currently working as Assistant Professor in Sree Chitra Thirunal College of Engineering, Pappanamcode, Thiruvananthapuram. Her areas of interest are network security, cloud computing and IOT.

Athira S Pillai is a final year B Tech student in the field of Computer Science and Engineering at SCT College of Engineering. Her areas of interest are data mining, machine learning and deep learning.

Disha Dinesh Majithia is a final year B Tech student in the field of Computer Science and Engineering at SCT College of Engineering. Her areas of interest are data mining, machine learning and deep learning.

