Query Processing using NLP

Nipun Hedaoo[1], Mohd Nomaan Khan[2], Mahesh Shetty[3], Ashwin Kutemate[4], Dr. Preeti Patil[5]

[1,2,3,4]B.E Students, Dept. of Information Technology, D.Y. Patil college Engineering, Akurdi, Pune, India
[5]Head of Dept., Dept. of Information Technology, D.Y. Patil college Engineering, Akurdi, Pune, India

---***---

Abstract – *The implementation of chatbots is widely increasing in all the sectors. It is a modern way to interact with computers to retrieve information. A chatbot allows a user to interact similarly as if, the user is conversing with another human. Firms from various sectors are implementing chatbots on their websites or applications to provide better customer service. The underlying technologies of a chatbot are Natural Language Processing (NLP) and Machine Learning. This paper successfully explained a chatbot that can be used to retrieve information related to Indian Railway such as Train Route, Train between Stations, Train Name/Number, Train Fare Enquiry, Train Arrivals, and Cancelled Trains, Rescheduled Trains.*

It provides a much more interactive way of fetching information while conversing with the Chatbot using voice and text in the mentioned methods. This proposed system will help common passengers and visually impaired people.

Key Words: Chatbot, Natural Language Processing, Machine Learning, Indian Railways

1. INTRODUCTION

Indian Railways is the fourth largest railway network in the world, transporting millions of passengers across the country. Over 23 million passengers travel daily in more than 12,617 trains. Similarly, a huge number of passengers use the web portal to get information such as booking train tickets, fetching PNR and train running status which tends to fail due to huge traffic. Even to get the above-mentioned information, People on railway stations stand in the long queues at the inquiry desk, which can be reduced by replacing a single information desk with multiple chatbots on every platform of the station.

India is a multilingual nation and the most common language used is Hindi. A majority of the crowd faces difficulty in conversing in the English language. Considering this issue, proposed system will be able to have basic conversation in English as well as in Hindi. The passengers using smartphones can save their time by using the Chatbot on their mobile phones.

2. LITERATURE REVIEW

Cyril Joe Baby, Faizan Ayyub Khan, Swathi J. N. in paper titled "Home Automation using IoT and a Chatbot using Natural Language Processing" discuss about a web-based application to control fans, lights and other electrical appliances over the internet. The web application hosts a chatbot to receive instructions from the user to control the functioning of the electrical appliances at home. The message sent using the chatbot is processed using Natural Language Processing techniques. The web application has a security feature that only enables certain users to access the application.

Chaitrali S. Kulkarni, Amruta U. Bhavsar, Savita R. Pingale, Prof. Satish S. Kumbhar in paper titled as "BANK CHAT BOT – An Intelligent Assistant System Using NLP and Machine Learning" discuss about a web application which deals with bank-related customer inquiry. The proposed system would help replicate the customer service experience at a bank resolving customer queries. The user's text format queries are converted into a vectorized format using vectorization. The classification algorithm is applied based on a trained dataset prepared from FAQs of banks websites that classifies the query and the result is given back to the user. Also, the web application has a feedback system that helps the user to report the given result by the bot.

Avni Prajapati, Pranali Naik, Sushmita Singh and, Anil Kale in paper titled as "Android Based Chatbot For College" discuss about an android application that provides the answer to any college queries of the students. The students would not have to visit the campus personally instead everything like notices, results, timetable, assignments would be made available at the fingertip just with the help of mobile phone and internet connection through a login. The students could easily retrieve information using a chatbot interface and any new data would get uploaded on chatbot's database by the admin.

3. REQUIREMENTS

The requirement for this project includes some software specifications which are: -

1. ChatterBot - ChatterBot is a python library used to develop interactive bots that can be trained in different languages to increase its applications. This system uses version 1.1.0.

2. SQL Alchemy - *SQLAlchemy* is the Python SQL toolkit. In this system, SQLAlchemy 1.2.0 is used to store the data in the SQLite database.

3. NLTK: Natural Language Toolkit for Natural Language Processing.

4. FLASK: FLASK for deploying the chatbot on a web-application.

5. Kivy: Kivy to deploy the application on different platforms.
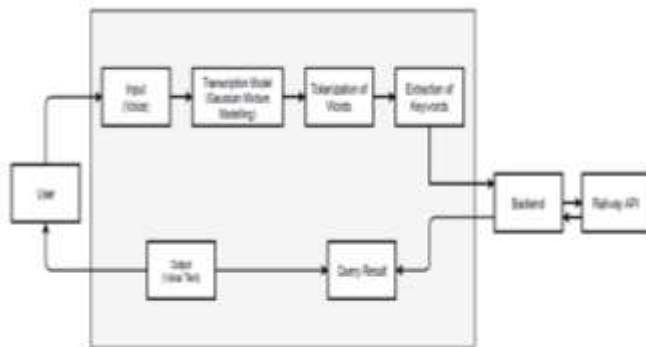
4. SYSTEM ARCHITECTURE

Figure 1 - System Architecture

5. PROPOSED SOLUTION

Most of the system which are currently being used as chatbot uses, 'Dialogflow' API. It is a natural language processing (NLP) platform that can be used to build conversational applications.

The proposed system will be based on python that can accept user input as text or voice. The system will have its own data set which includes answers of frequently asked questions. The data set also include greeting and basic conversation in English and Hindi language.

The system will use NLTK library for Natural Language Processing to decrease further processing and removing ambiguity and pre-process the query. Following steps are included in this task:

1. Removing punctuation marks and extra spaces.

2. Tokenization – The system will use tokenization to generate a sequence of words from the user's input query.

3. Removing stop words - Most of the common words like 'is', 'are', 'for', 'to' which can be ignored while processing the query will be removed.

Intent Classification: The system decides the intent of the user based on the query the user asks to the chatbot by recognizing relevant words. The intent indicates what information is required by the user like, PNR status, train running status, etc. The Railway API is called based on the intent of the user.

Processing relevant query: To retrieve the information in real-time from Indian railways the system will use the official Railway API provided by the Indian Railways. The railway API returns JSON objects from which all the information is extracted and displayed to the user on the chatbot interface.

The processed queries will be stored in the database and feedback from the user will be taken under consideration which can be used to improve the result of the system in future.

6. ALGORITHM

Algorithm used in this system are as follows:

Naïve Bayes: Naïve Bayes is a machine learning algorithm used for classification problems. This approach is primarily used for text classification, which involves high dimensional training data sets.

Using Bayes theorem posterior probability P(C|X) can be calculated as:

P(C|X) = (P(X|C) P(C))/P(X)

where,

P(C|X) = target class's posterior probability

P(X|C) = predictor class's probability

P(C) = class C's probability being true

P(X) = predictor's prior probability.

Recurrent Neural Network: Recurrent neural network is a class of nets that can predict the future based on provided data set. It can work on time series data and sequences of arbitrary lengths, rather than on fixed-sized inputs. Recurrent neural network is derived from feed forward network. Recurrent neural network looks similar to a feed forward network except it has connection pointing backward.

Decision Tree: Decision tree is the most powerful and most preferred tool for prediction and classification. A Decision tree is a flowchart similar to a tree structure. It is a map of all the possible outcomes of a series of related choices. In the tree structure each decision node indicates a decision to be made, each branch indicates a possible outcome or action, and each leaf node /terminal node indicates a final outcome.

SVM: SVM model is a representation of the data as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible.

Along with linear classification, SVMs can efficiently perform a non-linear classification. It can implicitly map the inputs into high-dimensional feature spaces.

7. CONCLUSION AND FUTURE SCOPE

Chatbots are capable of replacing humans at many workplaces and can help customers with accurate information in a shorter time period as compared to any other system. The proposed system in this paper is capable

of handling user queries related to Indian Railways and indulge users in basic conversation. It can replace Enquiry offices at railway stations in India.

In future, we are planning to add support to regional languages and improve the response time and accuracy of the system.

8. ACKNOWLEDGEMENT

We would like to express our sincere gratitude to our project guide Dr. Preeti Patil to let us work under her guidance and provide valuable inputs right from the beginning of the project. This project would not have been possible without the dedicated support of our guide.

9. REFERENCES

[1] Price Negotiator Chatbot Operating on Consumer Purchase Behaviour Mining Prof. Neha Pharande, Aditya Shukla, Ashish Shinde, Krishna Sikchi, Abhishek Kumar Singh, IJSRD - International Journal for Scientific Research & Development| Vol. 7, Issue 02, 2019 | ISSN (online): 2321-0613.

[2] BANK CHATBOT – An Intelligent Assistant System Using NLP and Machine Learning Chaitrali S. Kulkarni, Amruta U. Bhavsar, Savita R. Pingale, Prof. Satish S. Kumbhar. International Research Journal of Engineering and Technology (IRJET).

[3] Android Based Chatbot For College Avni Prajapati, Pranali Naik, Sushmita Singh, Anil Kale. International Journal of Scientific & Engineering Research Volume 9, Issue 4, April-2018 10 ISSN 2229-5518.

[4] Home Automation using IoT and a Chatbot using Natural Language Processing Cyril Joe Baby, Faizan Ayyub Khan, Swathi J. N. International Conference on Innovations in Power and Advanced Computing Technologies [i-PACT2017].

[5]. Daniel Toniuc and Adrian Groz- ―Climebot: an argumentative agent for climate change‖, Issue, 2017.

[6] Chatbot Evaluation and Database Expansion via Crowdsourcing, Zhou Yu, Ziyu Xu, Alan W Black, Alexander I. Rudnicky.

[7] http://www.nltk.org

[8] https://chatterbot.readthedocs.io/en/stable/

[9] https://railwayapi.com/api/