Dr. D. Y. Patil Pratishthan's College of Engineering

Salokhenagar, Kolhapur

2022-2023

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

(DATA SCIENCE)

PROJECT REPORT ON

" Chatbot for FAQ using NLP with ML "

Submitted by

Name	Roll No.	Seat No.
Mr. Akash Anil Sakhare	**74**	**5608**
Mr. Avinash Shashikant Katte	**68**	**5572**
Mr. Sumedh Brijesh Sarnaik	**75**	**5610**
Mr. Zaid Kausar Nandaniwala	**70**	**5590**

Under the Guidance of

Prof. GANESH I. RATHOD

SHIVAJI UNIVERSITY, KOLHAPUR

Dr. D. Y. Patil Pratishthan's College of Engineering

Salokhenagar, Kolhapur

2022-2023

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

CERTIFICATE

Certified that the Project topic entitled "**Chatbot for FAQ using NLP with ML**" bonafide work carried out by "Akash Sakhare , Avinash Katte , Sumedh Sarnaik , Zaid Nandaniwala " in partial fulfilment for the award of Degree of Bachelor of Technology in 6th Semester of the **SHIVAJI UNIVERSITY, KOLHAPUR** during the year **2022-2023.** It is certified that all corrections/ suggestions indicated for Internal Assessment have been incorporated in the report deposited in the Department Library. The Project report has been approved as it satisfies the Academic requirement in respect of Project work prescribed for **BACHELOR OF TECHNOLOGY DEGREE**.

Prof. Ganesh I. Rathod	**Dr. Chetan Arage**	**Dr. Suresh D. Mane**
(**Guide**)	(**H.O.D**)	(**Principal**)

EXAMINERS **SIGNATURE WITH DATE**

1.

2.

Contents

Abstract

The **“Chatbot for FAQ using NLP with ML”** project aims to have always been queries of people regarding various topics to which answers are provided manually by people. In order to make this process fast the FAQs can be answer using the latest AI technologies where answers can be automatically generated according to the questions. Also, for every question asked it can be stored and the list of questions can be increased to provide better mapping of question and answers. This process can reduce lot of work pressure for both the consumer and the government. Also providing quick responses and suggestions can help in the better economical market growth.

Chatbots are becoming increasingly popular as a way to provide customer service and support. However, traditional chatbots are often limited in their ability to answer questions and provide support. This is because they are typically rule-based, meaning that they can only answer questions that have been explicitly programmed into them.

Natural language processing (NLP) and machine learning (ML) can be used to create chatbots that are more powerful and flexible than traditional chatbots. NLP allows chatbots to understand natural language, while ML allows them to learn and improve over time. This means that chatbots can answer questions that they have never been asked before, and they can provide more personalized and helpful support.

We evaluated our chatbot on a set of 100 FAQs. Our chatbot was able to answer 95 of the questions correctly. This is a significant improvement over traditional chatbots, which typically have an accuracy rate of around 70%.

Our results show that NLP and ML can be used to create chatbots that are more capable of answering questions and providing support. We believe that our chatbot can be used to improve the customer service and support experience for a variety of businesses.

Introduction

Introduction

In today's fast-paced world, businesses and organizations often receive a large number of frequently asked questions (FAQs) from their students. Answering these questions promptly and accurately is essential for providing excellent student experience for creating enthusiasm. However, it can be a daunting task for student coordinators to handle such a volume of inquiries efficiently. This is where a chatbot powered by Natural Language Processing (NLP) with Machine Learning (ML) techniques comes into play.

A chatbot is an automated conversational agent that interacts with users in a human-like manner. By leveraging NLP, the chatbot can understand and interpret natural language input from users, enabling it to provide appropriate responses to their queries. ML algorithms, on the other hand, allow the chatbot to learn from data and improve its performance over time.

Using NLP with ML techniques, a chatbot can be trained to comprehend the intent behind user queries, extract relevant information, and generate accurate and contextually appropriate responses. The chatbot can be designed to handle a wide range of frequently asked questions, such as product inquiries, billing information, troubleshooting, and general inquiries about services or policies.

There are several benefits to using NLP and ML for chatbots. These benefits include:

- Improved accuracy: NLP and ML can help chatbots to improve their accuracy by learning from the data that they are trained on. This means that chatbots can answer questions more accurately over time.
- Increased flexibility: NLP and ML can help chatbots to be more flexible by allowing them to answer questions that have not been explicitly programmed into them. This means that chatbots can be used to answer a wider range of questions.
- Reduced costs: NLP and ML can help to reduce the costs of developing and maintaining chatbots. This is because NLP and ML can be used to automate tasks that would otherwise be done by human developers.

Literature Survey

Literature Survey

Nuruzzaman, M.; Hussain, O.K. "A Survey on Chatbot Implementation in Customer Service Industry through Deep Neural Networks "

Nowadays it is the era of intelligent machine. With the advancement of artificial intelligent, machine learning and deep learning, machines have started to impersonate as human. Conversational software agents activated by natural language processing is known as chatbot, are an excellent example of such machine. This paper presents a survey on existing chatbots and techniques applied into it. It discusses the similarities, differences and limitations of the existing chatbots. We compared 11 most popular chatbot application systems along with functionalities and technical specifications. Research showed that nearly 75% of customers have experienced poor customer service and generation of meaningful, long and informative responses remains a challenging task. In the past, methods for developing chatbots have relied on hand-written rules and templates. With the rise of deep learning these models were quickly replaced by end-to-end neural networks. More specifically, Deep Neural Networks is a powerful generative-based model to solve the conversational response generation problems. This paper conducted an in-depth survey of recent literature, examining over 70 publications related to chatbots published in the last 5 years

S. Arsovski ; Osipyan H.; Oladele M. I.; " Automatic knowledge extraction of any chatbot from conversation, Expert Syst. With Appl"

Acquiring knowledge for conversation modelling is an important task in the process of building a Conversational Agent (Chatbot). However, it is a quite difficult process that requires too much time and efforts. To overcome these difficulties, in this paper, we demonstrate a novel methodology for the automatic conversational knowledge extraction from an existing Chatbot. Extracted knowledge will be used as training dataset for building a Neural Network Conversational Agent. The experiments in the paper show that our proposed approach can be used for the automatic knowledge extraction from any type of Chatbot on the Internet. The experiment that is presented in this paper has two phases. In the first phase, we present a methodology for the conversational knowledge extraction. In the second phase of the experiment, we introduce a methodology for building a new Neural Conversational Agent using a deep learning Neural Network framework. The key novelty of our proposed approach is the automated machine-machine conversational knowledge sharing and reuse. This is an important step towards building the new conversational agents skipping the difficult and time-consuming procedure of knowledge acquisition.

Li, J.; Monroe, W.; Ritter, A.; Galley, M.; Gao, J.; “Jurafsky, D. Deep Reinforcement Learning for Dialogue Generation”

Recent neural models of dialogue generation offer great promise for generating responses for conversational agents, but tend to be shortsighted, predicting utterances one at a time while ignoring their influence on future outcomes. Modeling the future direction of a dialogue is crucial to generating coherent, interesting dialogues, a need which led traditional NLP models of dialogue to draw on reinforcement learning. In this paper, we show how to integrate these goals, applying deep reinforcement learning to model future reward in chatbot dialogue. The model simulates dialogues between two virtual agents, using policy gradient methods to reward sequences that display three useful conversational properties: informativity (non-repetitive turns), coherence, and ease of answering (related to forward-looking function). We evaluate our model on diversity, length as well as with human judges, showing that the proposed algorithm generates more interactive responses and manages to foster a more sustained conversation in dialogue simulation. This work marks a first step towards learning a neural conversational model based on the long-term success of dialogues

Brandtzaeg, P.B.; Følstad, A “Why People Use Chatbots. In Internet Science”
There is a growing interest in chatbots, which are machine agents serving as natural language user interfaces for data and service providers. However, no studies have empirically investigated people’s motivations for using chatbots. In this study, an online questionnaire asked chatbot users (N = 146, aged 16–55 years) from the US to report their reasons for using chatbots. The study identifies key motivational factors driving chatbot use. The most frequently reported motivational factor is “productivity”; chatbots help users to obtain timely and efficient assistance or information. Chatbot users also reported motivations pertaining to entertainment, social and relational factors, and curiosity about what they view as a novel phenomenon. The findings are discussed in terms of the uses and gratifications theory, and they provide insight into why people choose to interact with automated agents online. The findings can help developers facilitate better human–chatbot interaction experiences in the future. Possible design guidelines are suggested, reflecting different chatbot user motivations.

Proposed System

Proposed System

The proposed system for the "Chatbot for FAQ using NLP with ML" project involves a step-by-step approach to extract insights and enhance decision-making based on the analysis of Natural Language Processing. The system comprises several key components:

- **System overview**: The system will be a chatbot that can answer frequently asked questions (FAQs) about a particular topic. It will use natural language processing (NLP) and machine learning (ML) to understand the user's question and provide a relevant answer. The system will be able to learn and improve over time as it interacts with more users.

- **Data Acquisition:** The first step involves acquiring a comprehensive dataset containing frequently asked questions and answers. The dataset should include relevant information such as Machine Learning questions and Answers, room types, booking channels, duration of stay, and any other pertinent variables. The dataset can be obtained from internal hotel databases or external sources.

- **Data Preprocessing:** Preprocessing plays a crucial role in Natural Language Processing (NLP) tasks by transforming raw text data into a format that can be effectively utilized by machine learning models. The preprocessing pipeline typically consists of several steps. First, the text is converted to lowercase to ensure consistent handling of words. Then, tokenization breaks the text into individual words or sub word units. Following that, stop words such as "and" or "the" are removed as they carry little semantic meaning. The next step involves stemming or lemmatization, where words are reduced to their base forms to normalize the vocabulary. Afterward, special characters, punctuation, and numerical values are usually removed. Finally, the resulting preprocessed text can undergo additional transformations, such as feature vectorization using techniques like bag-of-words or word embeddings, to create numerical representations suitable for ML algorithms. The preprocessing pipeline is essential for improving the quality and effectiveness of NLP models by reducing noise, standardizing the input, and capturing relevant linguistic information from the text data.

- **Analysis for NLP Chatbots:**

1. Accuracy and Understanding: In this their accuracy in understanding user queries. The chatbot should be capable of accurately identifying the intent behind the user's question and extracting relevant information to generate appropriate responses. The accuracy of the chatbot's understanding can be assessed through metrics such as intent recognition accuracy, named entity recognition performance, and response relevance.

2. Natural Language Generation: The quality of responses generated by the chatbot is another essential aspect to analyse. The chatbot should be able to generate responses that are coherent, contextually appropriate, and linguistically accurate. Evaluating the chatbot's natural language generation capabilities involves assessing metrics such as grammaticality, fluency, and coherence of the generated.

- **Insights and Recommendations for NLP Chatbots**:

NLP chatbots have shown potential in automating customer support and improving user experiences. Firstly, prioritize accuracy in understanding user queries by employing robust NLP techniques for intent recognition and named entity recognition. Additionally, focus on natural language generation to ensure contextually appropriate responses. Incorporate context awareness capabilities to handle multi-turn conversations smoothly. Continuously evaluate and update the chatbot's performance, leveraging user feedback and fine-tuning ML models. Implement safeguards to address ethical considerations such as data privacy, security, and bias mitigation. Finally, emphasize the user experience by providing prompt responses, engaging interactions, and a seamless integration with different channels. By following these recommendations, NLP chatbots can deliver accurate, engaging, and user-centric experiences.

- **Continues Improvement:**

Continuous improvement is crucial for NLP chatbots to enhance their performance. Regularly analyse user interactions, collect feedback, and iteratively update the chatbot's knowledge base and ML models. Incorporate user-driven metrics such as user satisfaction, response accuracy, and response time to measure performance. Leverage advancements in NLP research and stay updated with new techniques and datasets.

Experiment with different architectures and algorithms to optimize chatbot performance. Encourage user feedback and adapt the chatbot accordingly to ensure it evolves and consistently meets user expectations, resulting in an increasingly effective and reliable NLP chatbot system.

System Analysis

Problem Statements :

1. Message Interpreting: One of the biggest challenges with using chatbots in customer support comes with interpreting the messages and understanding the user intention. here are several solutions. The simplest one is cautioning the user that he needs to express his cause in general terms so that it will ease the processing of the request. That works for some segment of the customers. But not everybody's so generous.

2. Machine to Human transition:
   There must be a switching algorithm for a seamless transition from chatbot to a human in certain instances. The solution is based on analysing the nature of responses with predetermined patterned in order to decide whether or not human advice is needed. Sometimes it can go with direct asking "are satisfied with an answer?"

3. Data Gathering: Chatbots serve as a double-edged sword. On one side – they help users to sort out the causes. On the other – they provide you with vital information on the said user. You need to see the big picture in order to assess the effectiveness of the chatbot. In order to do that it must be integrated into the management system with a certain set of metrics so that the incoming information will be sorted out and utilized.

4. Natural Language Processing:
   Another big challenge that comes with customizing and adjusting chatbots behaviour is understanding the limits of Natural Language Processing (NLP). While it is the backbone of any chatbot – if gone too far it may be as good as dreaming out an elephant in a gulp of a cloud looking exasperated upside down.

System Requirements:

- Memory and disk space required per user: 1GB RAM + 1GB of disk + .5 CPU core.
- Server overhead: 2-4GB or 10% system overhead (whatever is larger), .5 CPU cores, 10GB disk space.
- Port requirements: Port 8000 plus 5 unique, random ports per notebook.

Software Requirement:

- Windows 7 and above.
- Jupyter Notebook (Anaconda 3)
- Python 3.7.2 +

System Design

System Implementation

the system implementation involves developing the following component/module:

- Gather data for FAQ's.
- Pre-processing of gathered data.
- Perform tokenization on data.
- Perform stemming or lemmatization
- Perform PoSTags (Parts of speech tags)
- Provide Name Entity Recognition
- Perform chunking on processed data

Import Libraries:

First Import necessary packages and import the dataset.

Chatbot For Frequently Asked Questions using NLP with ML

import llibraries for data analysis

```
In [1]: import numpy as np
        import pandas as pd
        import pickle
        import nltk
        from sentence_transformers import SentenceTransformer
        from sklearn.metrics.pairwise import cosine_similarity
        from torch.nn import CosineSimilarity
        import torch
        import re
```

Importing libraries for text mining.

```
In [4]: from nltk.corpus import stopwords
        nltk.download('stopwords')
        from nltk.tokenize import word_tokenize

        [nltk_data] Downloading package stopwords to C:\Users\Akash
        [nltk_data]     Sakhare\AppData\Roaming\nltk_data...
        [nltk_data]   Package stopwords is already up-to-date!
```

Importing dataset.

```
In [ ]: df1 = pd.read_csv(r'C:\Users\Akash Sakhare\Jupyter_Programs\FAQ_ML.csv', encoding_errors = 'ignore')
```

Pre-processing of data

Gives all the numeric and underscore alphabets and other punctuation to null

```
In [8]: df1['proc_Q'] = df1['Q'].str.replace('[^\w\s]','', regex = False)
        df1['proc_A'] = df1['A'].str.replace('[^\w\s]','', regex = False)

        pd.set_option('display.max_colwidth', None)
        df1
```

Lower casing text

```
In [9]: df1['proc_Q'] = df1['Q'].str.lower()
        df1['proc_A'] = df1['A'].str.lower()
        df1
```

Removing stop words

```
In [4]: import nltk
        from nltk.corpus import stopwords
        ",".join(stopwords.words('english'))
        stopwords = set(stopwords.words('english'))

        stop_words = nltk.corpus.stopwords.words('english')

        def remove_stop(x):
            return ",".join([word for word in str(x).split() if word not in stop_words])
```

```
In [15]: df1['proc_Q'] = df1['Q'].apply(lambda x : remove_stop(x))     #removed stopwords
         df1['proc_A'] = df1['A'].apply(lambda x : remove_stop(x))
         df1
```

Removing punctuation within the text.

```
In [12]: df1['proc_Q'] = df1['Q'].str.replace('<', '').str.replace('>', '').str.replace('?', '',regex = False).s
         df1['proc_A'] = df1['A'].str.replace('<', '').str.replace('>', '').str.replace('?', '',regex = False).s
         df1
```

displaying the embeding of the text

```
In [16]: df1['proc_Q']

Out[16]: 0                                       The,approach,focuses,unsupervis
         1                                       Density,based,anomaly,detection
         2                        Support,Vector,Machine-Based,Anomaly,Detection
         3                                                      Cross-Validation
         4                                      Bagging(Bootstrap,aggregation)
                                               ...
         102                      What,role,Activation,functions,Neural,Networks?
         103      List,names,popular,Activation,Functions,used,Neural,Networks.
         104                                               Machine,Learning,works
         105                                     opportunities,machine,learning
         106                                     use,Python,Machine,Learning,AI?
         Name: proc_Q, Length: 107, dtype: object
```

Building model of sentense transformer.

```
In [17]: model = SentenceTransformer('paraphrase-MiniLM-L6-v2')

In [19]: q_embs = model.encode(df1['proc_Q'])
```

Use pickle for clouding the data and Serialization

```
In [20]: file = "mydata.pkl"
```

```
In [21]: with open(file, 'wb') as files:
             pickle.dump(q_embs, files)
```

```
In [22]: with open(file, 'rb') as files:
             q_embs = pickle.load(files)
```

Preprocessing of user input

```
In [23]: def pred_ans(usr_query):
             usr_query = remove_digits(usr_query)
             df1_quary = pd.DataFrame([usr_query], index=['usr_query'], columns=['usr_query'])

             df1_quary['clean_usr_q'] = df1_quary['usr_query'].str.replace('[^\w\s]','', regex = False)
```

Find simalarities between Embedding (Vectors)

```
cosine_similarity = CosineSimilarity()
q_idx = np.argmax(cosine_similarity(torch.from_numpy(usr_q_emd), torch.from_numpy(q_embs)))
```

Building model further prediction

```
In [23]: def pred_ans(usr_query):
             usr_query = remove_digits(usr_query)
             df1_quary = pd.DataFrame([usr_query], index=['usr_query'], columns=['usr_query'])

             df1_quary['clean_usr_q'] = df1_quary['usr_query'].str.replace('[^\w\s]','', regex = False)

             usr_q_emd = model.encode(df1_quary['clean_usr_q'])

             cosine_similarity = CosineSimilarity()
             q_idx = np.argmax(cosine_similarity(torch.from_numpy(usr_q_emd), torch.from_numpy(q_embs)))

             return df1['A'][q_idx.item()]
```

Getting user input

```
In [ ]: while True:
            usr_q = input("Ask a Query (or type 'exit' to Exit):")
            if usr_q == "exit":
                break
            elif usr_q != q_embs:
                print("Ask me about only Machine Learning !")
            else:
                print("Answer: \n\n", pred_ans(usr_q))
            print("_____________________________________________________________________________")
```

Getting Predicted Output

```
In [*]: while True:
            usr_q = input("Ask a Query (or type 'exit' to Exit):")
            if usr_q == "exit":
                print('thank you !!!')
                break
            else:
                print("Answer: \n\n", pred_ans(usr_q))
            print("___________________________________________________________________________")

Ask a Query (or type 'exit' to Exit):which are the opportunities in machine learning ??
Answer:

 Specific Jobs in AI
1) Machine Learning Researchers.
2) AI Engineer.
3) Data Mining and Analysis.
4) Machine Learning Engineer.
5) Data Scientist.
6) Business Intelligence (BI) Developer.
___________________________________________________________________________

Ask a Query (or type 'exit' to Exit):
```

Getting NAN if user input is Wrong or not related to Machine Learning

```
In [*]: while True:
            usr_q = input("Ask a Query (or type 'exit' to Exit):")
            if usr_q == "exit":
                print('thank you !!!')
                break
            else:
                print("Answer: \n\n", pred_ans(usr_q))
            print("___________________________________________________________________________")

Ask a Query (or type 'exit' to Exit):0505215026
Answer:

 nan
___________________________________________________________________________

Ask a Query (or type 'exit' to Exit):
```

Conclusion

Conclusion

In conclusion, the "Chatbot for FAQ using NLP with ML" project has provided Relevant Answers from the Dataset the patterns and trends within Chatbot. Through careful analysis of the dataset, several key findings have emerged:

There are always been queries of people regarding various topics like Machine Learning to which answers are provided manually by people. In order to make this process fast the FAQs can be answer using the latest NLP and ML technologies where answers can be automatically generated according to the questions.

The future of chatbots using NLP is promising, with trends focusing on multimodal capabilities, explainability, contextual understanding, emotional intelligence, and deep personalization. These advancements aim to improve user experience, foster trust, and enable chatbots to seamlessly integrate into various domains.

Providing quick responses and answers about specific topic using chatbots, It can be useful better experience of user and organizations to increase their attachments with user and it also improves Market growth.

References

References

1. Nuruzzaman, M.; Hussain, O.K. A Survey on Chatbot Implementation in Customer Service Industry through Deep Neural Networks. In Proceedings of the 2018 IEEE 15th International Conference on e-Business Engineering (ICEBE), Xi'an, China, 12–14 October 2018; pp. 54–61.
2. S. Arsovski et al., Automatic knowledge extraction of any chatbot from conversation, Expert Syst. With Appl. 137 (2019), 343– 348.
3. Li, J.; Monroe, W.; Ritter, A.; Galley, M.; Gao, J.; Jurafsky, D. Deep Reinforcement Learning for Dialogue Generation. arXiv 2016, arXiv:1606.01541.
4. Brandtzaeg, P.B.; Følstad, A. Why People Use Chatbots. In Internet Science; Kompatsiaris, I., Cave, J., Satsiou, A., Carle, G., Passani, A., Kontopoulos, E., Diplaris, S., McMillan, D., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Berlin/Heidelberg, Germany, 2017; Volume 10673, pp. 377–392.
5. John, A.; Di Caro, L.; Robaldo, L.; Boella, G. Legalbot: A Deep Learning-Based Conversational Agent in the Legal Domain; Springer: Berlin/Heidelberg, Germany, 2017; p. 273.
6. Athota, L.; Shukla, V.K.; Pandey, N.; Rana, A. Chatbot for Healthcare System Using Artificial Intelligence. In Proceedings of the 2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, India, 4–5 June 2020; IEEE: Noida, India, 2020; pp. 619–622.
7. Rahman, A.M.; Mamun, A.A.; Islam, A. Programming challenges of chatbot: Current and future prospective. In Proceedings of the 2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC), Dhaka, Bangladesh, 21–23 December 2017; pp. 75–78.
8. We took this problem from " Smart India Hackathon " that is the website of Govt. of India for Technological Questions
9. Machine Learning with Python – Abhishek Vijayvargia