

Python Chatbot Project Report

Revised on 31st Mar 2023

PYTHON CHATBOT PROJECT

NLP TECHNIQUES

Instructor: Dr. Karen Mazidi

Prepared for:

Natural language Processing Techniques

Prepared by:

Michael Muruthi

Table of Contents

1.	Introduction	3
1.1	Purpose	3
1.2.	Scope	3
1.3	How it works	3
2.	NLP Techniques Used	4
2.1	Sentiment Analysis	4
2.2	Vectorization	4
2.3	Training	5
3.	Conversation Model	5
3.1	Dialog Tree	5
3.2	Sample Dialog Interaction	6
4.	Chat bot Evaluation	8
4.1	Strengths of the Chat bot	8
4.2	Weaknesses	8
4.3	User Reviews	9
4.4	Knowledge Base Appendix	10
4.5	Appendix User Models	11

1. Introduction

The need to have automated language processors has been increasing by the day. Many organizations are turning to electronic methods of serving their customers, a trend which is anticipated to continue and change the way many organizations work in the future.

1.1 Purpose

The purpose of this project is to develop a chat bot for a car vending site to enable customers to obtain details about the prices and availability of cars without the need for human interaction.

Essentially, it involves the implementation of a virtual customer assistant at a car seller's site, which facilitates prospective buyers in obtaining information about available car models and prices. This chatbot functions by receiving user input, processing it, and crafting an appropriate response using already existing data from a knowledge base scraped from the internet.

1.2. Scope

The chatbot is limited to the data available on the site's knowledge base and therefore may not be intelligent enough to serve with information outside the given scope.

1.3 How it works.

When a user initializes an instance of the bot program, some initial data processing is done. First, the knowledge base is loaded and transformed to actionable data. The user is then prompted to begin the conversation with a greeting line and a prompt for a name. Every user statement is evaluated against predefined phrases and analyzed using NLTK and sklearn tools to determine an appropriate response. The process of analyzing involves, removal of stop words, punctuation marks and stemming which creates a list of words. Further processing creates classifications of words based on occurrence of each individual word, within the knowledge, a process called vectorization.

Some predefined phrases are also defined in the program to help in generation of responses. This includes greeting texts, exit messages and prompt statements, as well as some appropriate responses.

To ensure that the bot can work better in future, some personal information is gathered during the conversations. When a user indicates dissatisfaction with a response, the subject is recorded to ensure that it is not displayed in future.

2. NLP Techniques Used

To ensure that the bot responds better several Natural Language Processing tools were used. These tools include.

2.1 Sentiment Analysis

Sentiment analysis is a Natural language processing technique which involves the analysis of sections of sentences. The analysis is meant to determine the intent, subjects, tone, and any domain specific values.

The chatbot uses NLTK module, a python module which offers many sentiment analysis tools, to analyze user input statements and determine the intent of the user. Some of the NLTK tools that are used in the project include:

1. Stop Words -> (Words which are regarded as to be of less effect to the meaning of a given phrase)
2. Lemmatization -> (Tool aims at removing inflectional endings only and to return the base or dictionary form of a word)
3. Stemming -> (This is the reducing of inflected words to their word stem, base or root form)

All these combined helps extract the intent of a given user.

2.2 Vectorization

When the intent of the user has been determined, an appropriate response must be displayed. To ensure that the response is relevant, vectorization is applied to determine the weights of responses in relation to the user intent. This is done using term frequency-inverse document frequency (TF-IDF). TF-IDF weight

is a statistical measure used to evaluate how important a word is to a document in a collection or corpus. The importance increases proportionally to the number of times a word appears in the document but is offset by the frequency of the word in the corpus.

This is achieved by measuring term frequencies and vectorization modules available in Python's sklearn module. These two helps in determining the relevance of a response in regard to a user input.

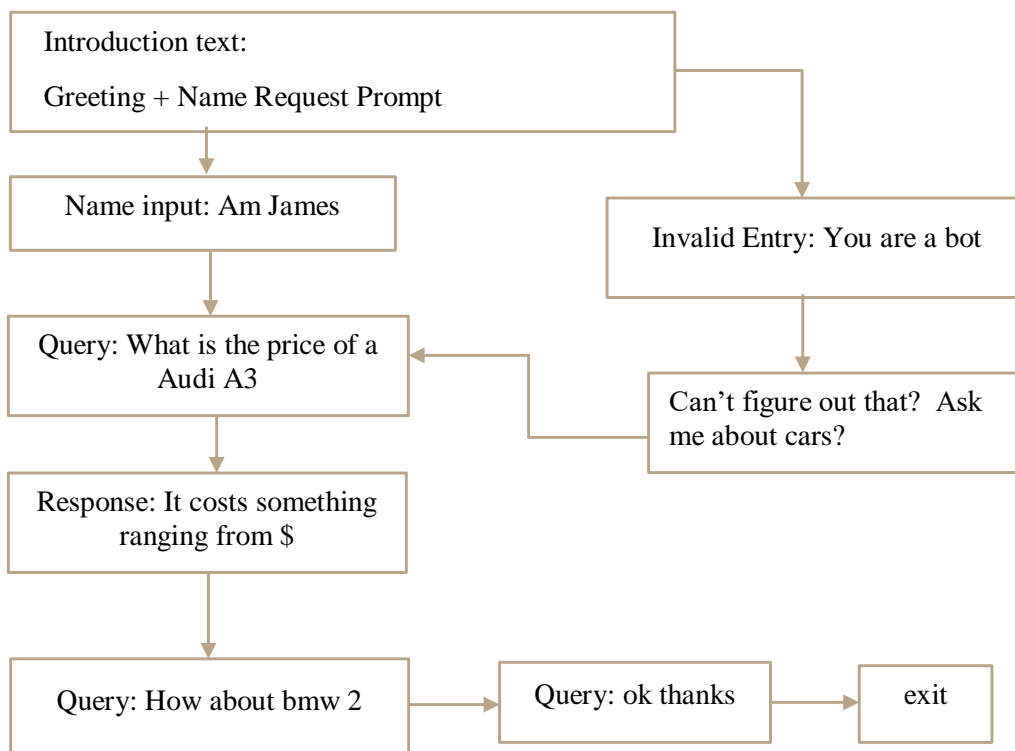
2.3 Training

As the conversations are being carried out, some data is gathered by the bot to help in generating better responses in subsequent conversations. In this case, the bot keeps a record of the name of the user and a list of the likes and dislikes gathered during the conversation. By this the bot can display better responses and more relevant answers.

3. Conversation Model

The chat bot's conversation model is based on the user input. The user input is evaluated, and responses are generated. Here is a dialog tree for a sample conversation.

3.1 Dialog Tree



3.2 Sample Dialog Interaction

```
How are you doing
What is your name?
User:>>Am Tonny
Your name is Am Tonny Right?
User:>>no
So what is your real name?
User:>>Tonny
You are Tonny Right?
User:>>yes
Welcome Tonny
Which car are you looking for?
```

```
Tonny :>>Am looking for an audi A3
a 2018 audi a3,hatch has a price or cost starting from $42702, audi
a3,sedan has a price or cost starting from $49012.
:>> If that is irrelevant, please elaborate :-( or just type the
name of another car
Tonny :>>Wow what about a audi convertible
audi a5,convertible has a price or cost starting from $94924.
By the way sedans are cheaper!
Tonny :>>Fine finally a bmw x series
bmw 4 series,coupe has a price or cost starting from $78829, bmw 4
series gran,coupe has a price or cost starting from $78829, bmw 4
series,convertible has a price or cost starting from $99166.
:>> If that is irrelevant, please elaborate :-( or just type the
name of another car
Tonny :>>ok thanks
ChatBot :>> Thanks for your time. Bye
```

3.2 Elaborate sample dialog Interactions

```
===== RESTART: F:\PycharmProjects\advchatbot\chatbot.py =====
```

```
Hey!
Welcome, Please enter your first Name?
User:>> Michael
You are Michael Right?
User:>> correct
Welcome Michael
Enter a vehicle make and model.

Michael :>> honda accord
the honda accord boasts a nice engine and transmission that allow you to ride comfortably down the road
honda is a well-known brand and many consumers consider it one of the most reliable too
with that said, a honda accord can easily rake in 200,000 miles or more without any major mechanical issues under the hood-
no matter the year of the used honda accord you are eyeing, you will be in good hands.
Michael :>> are there great features in a honda accord
the honda accord boasts a nice engine and transmission that allow you to ride comfortably down the road
honda is a well-known brand and many consumers consider it one of the most reliable too
with that said, a honda accord can easily rake in 200,000 miles or more without any major mechanical issues under the hood-
no matter the year of the used honda accord you are eyeing, you will be in good hands.
```

```

:>> Keep rolling, we have numerous vehicles. Type another vehicle name!
Michael :>> What is the best way to purchase a vehicle
the best way for you to purchase a vehicle is to do so face to face in a public space and never exchange money for the car unless you feel comfortable doing so.
:>> try body styles also such as sedan, suv, coupe, hatchback, convertible etc
Michael :>> What are the pros and cons of dealer
pros and cons of dealer purchase
    pro: the dealer is responsible for handling all the paperwork and the process
    pro: you have financing options you can choose from
    pro: you have the option of both certified and non-certified used cars
    con: you may purchase a used car that is more than your budget allowed for
    con: you may feel obligated or pressured into a buy this car situation
    con: used cars are often higher priced.
:>> Keep rolling, we have numerous vehicles. Type another vehicle name!
Michael :>> What are some of the safety regulations I should know
there are two main sources for you to obtain the safety information you need to about the used car for sale you are eyeing: first, you can run a check for safety ratings at the insurance institute for highway safety or the iihs- the second source is the national highway traffic safety administration or the nhtsa; both sources will provide you with a better idea of how safe the used car you are interested in is and how well it performed in any crash tests.
By the way, sedans are cheaper!
Michael :>> What features does a honda civic have?
honda civic suv has a price or cost starting from $19835 the civic exemplifies automotive excellence and mixes fun with efficiency and practicality.
    :>> If that is irrelevant, please elaborate :-( or just type the name of another car
Michael :>> What is the safety score for a nissan altima?
Spell Check: >> Corrected Words
['ultima']
safety is paramount in any car and the 2017 nissan altima is a supreme safety machine- in fact, the iihs scored it as good in all areas of crashworthiness and ranked it superior for front crash prevention - with an overall five out of five for safety ratings, it is no wonder that individuals and families choose a used 2017 altima over other options.
:>> try body styles also such as sedan, suv, coupe, hatchback, convertible etc
Michael :>> Tell me about car scams and what to lookout for?
buying a used car is a time of happiness, excitement, and joy, but, it is also a time for you to make sure that you are cautious and paying attention too- used car scams do exist and you must be on the look out for some of the most common scams or you may find yourself stripped of your money and stuck with a car that doesn't work.
    :>> If that is irrelevant, please elaborate :-( or just type the name of another car
Michael :>> Which cars are the most reliable?
as you shop for a used car, do take the time to do your homework about the car- one thing you want to keep in mind is that some cars are just simply troublesome from the beginning and there are some brands and models that are less reliable than others- to help limit how many repairs you will need to make on your used car, search for one that is reliable.
:>> try body styles also such as sedan, suv, coupe, hatchback, convertible etc
Michael :>> What are used cars in particular?
now that they have you at the dealership, they can sell you other, higher-priced used cars.
    :>> If that is irrelevant, please elaborate :-( or just type the name of another car
Michael :>> What does it mean by certified cars?
certified used cars generally come with a warranty attached to them and a dealer can back up any claims made that the car is certified- it is important that as you shop for a certified used car that you do not just choose a car because there is a certified sticker on it.
    :>> If that is irrelevant, please elaborate :-( or just type the name of another car
Michael :>> How are the toyota prius?
Spell Check: >> Corrected Words
['pius']
the toyota prius took the automotive world by storm by offering some of the best fuel economy to consumers- the car was manufactured with technology and simplicity in mind- each prius can drive on the road using only electricity up to about 25 miles per hour, which helps reduce the cost of ownership too.
:>> try body styles also such as sedan, suv, coupe, hatchback, convertible etc
Michael :>> bye
ChatBot :>> Thanks for your time. Bye
>>>

```

4. Chat bot Evaluation

4.1 Strengths of the Chat bot

1. Can be trained

The user model makes it easy to scale down to a user preference as the user interacts with the system. As a user interacts with the bot, a user model is updated with the user preferences thus indirectly training the bot on how to respond better.

2. Automatic

The bot interacts with people just like they would interact with other human beings. This makes it a good option for customers seeking general information or customer service during the day and also during the night or offline hours without having a person to monitor and respond to conversations.

3. Faster access

The Chat bot can access information faster than human beings and can give suggestions better than humans, because of its processing nature.

4. Serving Many people at the same time

Being a software system, a chat bot can be instantiated to serve many people at the same time effectively.

4.2 Weaknesses

1. Limited Domain

The bot can only serve people on a limited number of user queries. This is because the knowledge it acts on is based on a particular topic; in this case car prices.

2. Not perfect

As much as data can be supplied and user models created, the bot cannot be perfect. It still cannot make decisions as human beings can. Sometimes it may respond with accurate but irrelevant information. In cases where the use references previous conversations for other days, it may not remember like a human being.

3. Require a user to reframe sentences

In case a sentence is complex, the chat bot may not be able to determine the right cause of actions. In a phrase where multiple intents have been raised, the bot may not be able to pick an appropriate response. On a different note: in a case where a user mixes both negative and positive expressions of the same intent, it may be difficult to determine what they exactly meant.

4.3 User Reviews

After creating the bot, I had three 3 coworker test and it and gave the following reviews:

The survey question was:

“On a scale of 1 to 10 with 1 being worst and ten the best, what would be your rating? Give a description of the strengths and weaknesses in a paragraph”

1. User 1: Andy

Well, I would give it a 9.

First it is a great idea and works nicely. I like that got my name well. I got the prices of different cars but some other models are missing! Adding more data would make it great. It understands well when the conversation is over and types an exit message. However, at sometimes it gives irrelevant responses! This is why I gave it a nine.

2. User 2

Men that was great: an 8 for it.

I got many prices of many cars, but at a certain point I got a slightly irrelevant response. If you can improve on the responses.

3. User 3.

I will give an eight

Good work and great idea. There are prices for a number of cars, but it cannot remember when I ask of a previous query! It keeps telling me that it does not understand!

4.4 Knowledge Base Appendix

The knowledge base contains data scraped from a website where they deal with cars reviews.

Below is a sample of the data from the knowledgebase.

If safety is one of your main concerns in a vehicle, you will find that this truck stacks up. It received a four out of five stars in overall safety. The Tacoma also received a five out of five stars for side impact crashes as well- According to the IIHS, the 2015 Tacoma received a testing of good in the crash areas of head restraints and seats, side, and front- The lowest rating, which was marginal, was roof strength; however, the Tacoma did score a four out of five in rollover crashes for safety.

Ford Focus,Sedan has a price or cost starting from \$28347. Ford Mondeo,Hatch has a price or cost starting from \$39490

The Tacoma has earned its place in the truck lineup and continues to make consumers happy with their choice to purchase one; Consumers who do own this truck has the following to say about it:

The average price of a used 2015 Toyota Tacoma is somewhere between \$15,000 and \$20,000. This truck is known to hold its value over time, which makes it a great investment.

Ford Mustang,Coupe has a price or cost starting from \$69970. Ford Mustang,Convertible has a price or cost starting from \$83117

4.5 Appendix User Models

While a user is interacting with the bot, it gathers information into a user model. The model is identified by the user name and contains the likes and dislikes of the user.

Here is a sample user model

```
{ UserName: John:{  
likes{[sedan, bmw,audi, Volkswagen,hatch]},Dislike{['Mercedes', 'Mitsubishi']}}}
```

References.

Knowledgebase:

First Url -> <http://thefactfile.org/car-facts/>

Second URL - > <https://www.alautotransport.com/container-car-shipping-faq-and-answers/>

Third URL -> <https://www.consumerreports.org/car-reliability/10-most-reliable-cars/>