Indian Institute of Technology, Palakkad

Internship Report

Appzillon Chatbot and Voicebanking Service in Python

---

by

S S T Siddhardha
111501028
Dept. of Computer Science and engineering

Project Mentors
Balachandra R
Sundarrajan R

Place of Internship:
I-exceed Technology Solutions Pvt Ltd,
SJR Padukone Towers,51, 100 Feet Rd,
2nd Block, Koramangala, Bengaluru

Period of Internship:
2-May-2018 to 11-July-2018

Mentor Signature

Preface

This report presents the details and experiences of my internship done at i-exceed technology solutions private limited.During the period of internship I worked on building the chat-bot for applications in Banking Sector.

This report mainly addresses the different stages in building chat-bot like Model Generation, Prediction, Auto Suggest and algorithms used in respective stages.

Acknowledgments

Firstly I would like to express my sincere gratitude to Ushant Mukharjee of i-exceed for giving me the internship. I would like to express my indebtedness appreciation to Balachandra and Sundarrajan of i-exceed for constant guidance and advice during internship.I am grateful to my colleagues during internship Athul and Jayaprakash of IIT Palakkad. Lastly I would like to thank my department Internship Co-coordinator and Training and Placements Officer.

About Company

I-exceed is a digital transformation partner for leading financial institutions worldwide. Their offerings include Appzillon Digital Banking, a suite of pre-built omni-channel digital banking solutions and Appzillon Digital Platform, an award winning low-code application development platform that facilitates rapid digital transformation.

People at i-exceed divide in to two streams, One part of people work on developing the software Appzillon which is used for building banking applications and other part on creating apps using Appzillon for Banks.Some of the industry leaders they work are Andhra bank, Canara Bank, Karnataka Bank..

Abstract

This report refers to the work completed during my internship with i-exceed technology solutions private limited. It present the three main tasks done during summer internship which are listed below

i Model Generation and Prediction

ii Auto Suggest

iii Adding Voice Layer

All these tasks have been completed successfully and results were according to expectations. Model Generation and Prediction was done using three different algorithms which discussed in details in the report and best one was selected. Auto Suggest was done using Markov Model. Adding Voice layer to text based Chat-bot was done with the help of third-party software named Microsoft Azure Speech to text.

Contents

1 Introduction

Chatbots are at the talking point of the business world now. A Chatbot ("Chatter Robot") is a computer program that mimics human conversations in its natural format including text or spoken language using artificial intelligence techniques such as Natural Language Processing (NLP), image and video processing, and audio analysis. Chatbots can schedule meetings, tell you weather forecast, provide customer support and many more.

Tasks involved in building Chatbot are listed below.

1 Dataset Generation and Feature selection

2 Model Generation

3 Prediction and improving results.

Apart from these for text based Chatbot we can add following.

1 Auto-Correct or Spell-Checker

2 Auto-Suggest or Auto-Complete

Voice based chatbot can be made by just adding a layer over the existing text based chatbot. Tasks involved in building Voice based chatbot is extracting text from speech, finding required information from extracted text, asking the user to provide necessary information.

I used python language for programming and nltk module for training the Model.

2 Model Generation

The Chatbot we built is for Banking applications, Sample sentence that user communicates looks like following

- Open account for Google in Rome.
- Create deal for Microsoft at Lux.

We need to identify three entity types from each transaction/sentence.They are **Clientname, Product, Branch**.
Lets take a look at what these entity values will be for sentence *Open account for Google in Rome.*

- Clientname : ***Google***.
- Product : ***Open account***.
- Branch : ***Rome***.

2.1 Generating Training Data

Training file for entity type **branch** looks like
Open account for Google in <START: BRANCH> Rome <END>

So for each entitype we need to seperate data. For ex. In clientname model we need to expand other tags and keep in clientname tag intact.
Sample utternamce : Open <START:product> %product% <END> for <START:clientname> %clientname% <END> in <START:branch> %branch% <END>.
Generate all possible sentences by expanding the other tags (product, clientname in this case) with all product,clientname entities. This is done for all the utterances. Preprocessing of data like removing stopwords, converting to lower case is done.

2.2 Feature Generation

Now we need features for each word and its label for training.
Eg: Open account for Google in Rome
word = 'Google'
Features = {previous word : for; previous previous word : account; next word = in; next next word : Rome; wordcase : lowercase; prev.wordcase : lowercase; next.wordcase : lowercase;...}
Label = clientname-start

When using nltk we need to give features for each word and its label so we have freedom to select the features whereas in the case of OpenNLP it automatically chooses the features once label is specified.

2.3 Model Generation

Now train the model using three different Algorithms Naive bayes, Maxent, Decision Tree.When i tried with 4:1 (train:test) data using naivebayes i got around 70% accuracy but my colleague got more than 80% so he continued with improvement on that model.
The reason i found for getting less accuracy is using less features.POS tagging is a important feature which i missed.

3 Prediction

When user enters a sentence we need to predict the clientname, product and branch from the sentence.For that we need to generate features of each word in the sentence and find out the label for each word using the trained models.

3.1 Sample Case

Lets consider a case where user enters **Open account for SSTS in Amaravathi**

User can specify the algorithm to use for prediction. But there is default algorithm set when user don't want to mention the algorithm. Before Predicting we need to generate features of each word in above sentence and predict label from using previously trained models (One model per each entitytype). Final Output Labels foe each word in sentence looks like

- Clientname : ***[SSTS]***.
- Product : ***[Open account]***.
- Branch : ***[Amaravathi]***.
- Other : ***[for, in]***

3.2 Processing text from screen to backend

The front-end screen was made using HTML, CSS and javascript. The ext users enters is sent for prediction through Ajax request. In the python program(predictor service) the text entered is received and cross validate the values of each Fields entered and processed to predict the labels for each word, then output is sent back to HTML screen.

4 Autosuggest

Autosuggest is a feature which an application preidcts the rest of word/sentence user is typing. In graphical user interface users can typically press the tab key to accept a suggestion or the down arrow key to accept one of several.
Advantages of Autosuggestion are

- Reduce keystrokes.
- Human Computer interaction will be faster.
- help avoid typos.

4.1 How it works

Method i use here is related to Markov Model.

Given an input string that consists of N words w_1, ..., w_N, the model predicts the following word, w_N+1, from the language model.The most probable candidate for w_N+1 is computed by maximizing P (w_N+1 | w_N, ..., w_N - O + 2) ,where O is order of model.

For instance , consider a 2-gram model and sentence w1,w2 w3 is chosen such that P(w3 | w2) is maximized.

Higher-order model will be more precise, but at the expense of generating a large list of n-grams, which may negatively impact on storage space and computational time.

Problems I encountered at this stage are dealing with case sensitive, drwaback contect can also be used

4.2 Training the data and Predicting next words

Firstly we check for whether model exists or not.If not then compute language models for all values of ngrams(1 to N).Generate ngrams for each sentence for each value of 1,2,....N.
Let W1,W2,...Wm be words entered so far by the user now we will check for all the sentences in the training data starts-with W(m-O)....,Wm-1,Wm . Now calculate the probability for each sentence and find the most 10 probable sentences and list them in drop down list.

5 Voice Layer

The big difference between voice based chatbot and text based chatbot is how we interact with them.A user interacts with a voice based chatbot differently: they converse with such a bot via their voice in natural language.A voice chatbot enables users to accomplish tasks on these devices hands-free. A messenger chatbot might be more tempting to users because they can easily begin chatting with it on their phone. Unless voice chatbots exists on a user's phone or computer, interacting with it requires buying a new smart speaker device. That said, voice activated chatbots are the ideal interface when hands-free interaction is necessary, like walking the user step-by-step through a recipe when their hands are full. Voice chatbots can also be used by blind people.

We used already built text based chatbot and added a voicelayer above it to behave as voice chatbot.Task involved in building it are :

- Capturing the speech.
- Converting speech to text.
- Getting required information from user.

5.1 Converting speech to text

There are many third party softwares available for converting speech to text.We can also write a program for speech recognition using packages like tensorflow(uses neural networks).But for that we need large amount of dataset.Since the company we work is into banking sector they don't want to use general dataset that is used for speech recognition.So generating dataset in this case is a problem.Though research in this area is done by my colleague, we are guided to use the third party software Microsoft Azure Speech to Text.

Getting required information from the user is most import thing.Lets say there are n fields that are to be entered user may wish to speak only once with all values for all the fields at once or he may wish to speak value of each filed at once. We need to recognize values for a particular filed and ask him to speak about unanswered fields. For this we need a Dialogue Tree.

5.2 Extracting information from text

For extracting information from text we need to have separate models for each filed in the screen.For that we need to have training data. We generated a small set of training data for each filed like Algorithm to use in prediction, and generated model like we did in previous chapters. After extracting information from text we cross checked the extracted information with possible values for respective fields and updated flags for those fields which are to be used in next part ie.. Dialogue Tree.

5.3 Dialogue Tree

Dialogue Tree is a way of letting user know what are fields that user needs to enter next.We used Logistic Regression for building dialogue tree.
Suppose there are N fields that are to be entered then training data for dialogue tree will be a list of N flags (zeros and ones) where each answered have field have respective flag as 1 and unanswered filed have respective flag as 0 and label(lies in between 1 to N that indicates the next question to be asked.

Now when user speaks we try to extract values for all the fields and change the flags corresponding to answered fields and predict the next question to be answered using previously build dialogue tree.

Problems we encountered in the building the voice layer are Speech recognition was very bad for some third party softwares, that is because we think that banking terminology is not so common to get good results.

Text to Speech part is not implemented because of security reasons bank generally don't want that.For instant if user asks for balance , user don't want his balance to be revealed to third party.

6 Summary

Banking sector already emerged as important vehicle for business. Now a days AI is almost entering every department in human life.Linking Banking sector with AI is a good idea and also favours fast and easily accessible transactions. Things i learned during my internship were writing a clean code so that others can understand, i learnt to differentiate between things what industry expects and what we generally do, importance of teamwork. Technically i learnt many things like using nltk, maxent algorithm...

7 Future Scope

Incase of training the models we can reduce the size of dataset.Currently we are producing all combination of entities for a single utterance, there is only change of name but not sentence structure.we can reduce that.Instead of having three different models for three entity types , we can have only one model for all the entity types.For that we need large data.Incase of Autosuggest we can take note of sentences that user entered and periodically train the models for better results(suggestions). For speech to text extraction instead of using third party software we can write a program using tensor flow,but it requires large dataset.