

SOFI: Your Personal Assistant

Aakash Dixit¹, Amandeep Singh Oberoi², Anshul Jaiswal³, Kushal Agarwal⁴

Students, CSE Dept. of Guru Gobind Singh Indraprastha University
Dwarka New Delhi

Sharanya Chandran⁵

Assistant Professor CSE Dept. HMRITM

Abstract - The exigency to make our lives easier and automating our quotidian tasks, virtual assistants are heavily relied upon in today's bustling world. Sofi assistant depends upon speech to text recognition and use of intent classifiers to predict the underlying objective and produce a productive results which meets the end goal of the user. Through the use of web scrappers, automation, recommendation models and many more we are successful in producing a fruitful outcome. The main idea behind the project was to build a cost effective project with minimal processing and fast results to achieve end intent of the user. It is used to target the social media, everyday activities like news, weather and lot of other queries to garner stimulating results. Future scope of the project would include to accept any mode of customer input, example-text, speech, video, image as well as the integration of home automation combined with the use of IoT.

Key Words: Voice assistant, Neural Network, Intelligence, Speech Recognizer, Internet,

1. INTRODUCTION

In earlier time people just want to have a genie who can act over command and get their work done. Isn't it amazing that, in today's world you can get your work done from a machine just by your voice. In 21st century, everything is leaning towards automation, may it be your phone or laptop or home or car. Our lives have become more fast-paced and hectic than ever before! We communicate, interact and learn online. An easy access to machine with voice commands is the revolutionary way of human system interaction. To achieve this, we need to use speech to text API for understanding the input. Thus we have developed SOFI Personal Voice Assistant which depends upon speech to text recognition and use of intent classifiers to predict the underlying objective and produce a productive results which meets the end goal of the user. Having the ability to interact with the surroundings just by one of the materialistic form of human interaction i.e. HUMAN VOICE, this application is of great use for Blind people who can't access their PC with the help of keyboard. It will also make work

easy and efficient for Normal Users. This application can Run on any Operating system. So there arise a question like how to use it? This device can be very handy for day to day use and it can help you function better by constantly giving you reminders and suggestions. You just need to say Open Whatsapp or any other application and it will do the same for you. Because your own voice is turning into a best input device than a conventional enter key, the working of application is Voice Controlled, which will return answer to your query in voice as well as text form.

1.1 Meaning Of Voice Assistant

In general voice assistants react to voice commands and give the user relevant information about his inquiry. Presently voice assistants are already able to process orders of products, answer questions, perform actions like playing music or start a simple phone call with a friend. The basics of the technology currently exist and the next few years will be used to develop these artificially intelligent assistants even further, enabling them to have more complex capabilities. The long-term vision for voice assistants is to act as a smart bridge between humans and the vast knowledge and capacities which the internet delivers. Taking away the need to use any device or screen to interact with the internet, technology or other humans in different locations

1.2 Working Of Voice Assistant

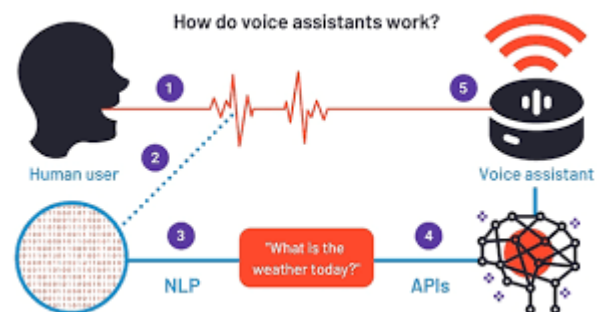


Fig-1 : Working Of Voice Assistant

For the voice-enabled world so-called smartspeakers have shown to hold a lot of potential. These have a microphone to

“hear” and speakers to communicate back to us or play music. The smart part is their direct connection to the internet and advanced speech recognition software. The software can understand and react to specific commands. Asking the voice assistant in the smart speaker for tomorrow’s weather is a good example. The assistant will decipher the human voice command and then act independently by searching the web for the right answer and reading it back to the user. The smart speaker itself is only the physical part of the smart assistant, the connecting device between bits and bytes and our human needs and demands. The software is the actual core of the artificial intelligence that enables the smart speaker to answer our specific questions and human commands. Being able to understand and process our questions is the great challenge voice assistants have to master. Meaning a voice assistant has to be able to learn and make use of languages to be able to pick the right answers to our commands. In the combination of the physical and digital parts of the smart speaker lies their great potential to act as a bridge in-between humans and their computers or machines. Recent studies show that the rate of adoption of these new smart devices is even faster than it was with the internet or mobile, as humans are getting more used to technology being a consistent part of our lives.

2. Literature Survey

Speech recognition has a long history with several waves of major innovations. Speech recognition for dictation, search, and voice commands has become a standard feature on smartphones and wearable devices. Design of a compact large vocabulary speech recognition system that can run efficiently on mobile devices, accurately and with low latency.

This is achieved by using a CTC based LSTM acoustic model which predicts context independent phones and is compressed to a tenth of its original size using a combination of SVD-based compression and quantization. Quantized deep neural networks (DNNs) and on-the-fly language model rescoring to achieve real-time performance on modern smart phones. The ASR and Search components perform speech recognition and search tasks. In addition to ASR and Search, we also integrate a query parsing module between ASR and Search for a number of reasons. Set of techniques for improving the performance of automated voice search services intended for mobile users accessing these services over a range of portable devices. Voice search is implemented as a two stage search procedure where string candidates generated by an automatic speech recognition (ASR) system are re-scored in order to identify the best matching entry from a potentially very large application specific database. Study provides a good example of how additional domain specific knowledge sources can be used with a domain independent ASR system to facilitate voice access to online search indices. As more data becomes available for a given speech recognition task, the natural way to improve recognition accuracy is to train larger acoustic models. There are a nonparametric empirical model that exploits abundant

training data to directly learn pronunciation variation. Interpolating the empirical model with a parametric model yields the best performance, with a relative improvement of 5.2% in WER over the baseline. There are a number of ways in which this work could be extended. First, closer integration with acoustic model training is likely to yield sharper distributions and a tighter fit to the data. Second, estimating word pronunciation co-occurrence counts in semi-supervised fashion (e.g. through word recognition instead of forced alignment) would broaden its applicability to a wide range of speech genres and tasks.

3. Proposed Architecture

SOFI has various branches of the services, but the main feature of SOFI is Voice Recognition Engine which has an ability to work without internet connection i.e. Offline Voice Recognition. This Voice Recognizer works offline and performs various operations as per the user commands and requirements. This is the first activity that’s opens whenever we starts SOFI and it directly can be opened by pressing SOFI App Icon on the Desktop. This feature is specially design for Blind Persons who wish to use the computers but are unable to connect this technology. Also Native user who barely knows to Switch ON their computers can easily open this application and using voice commands in their local languages as per need. SOFI responds to basic commands like, Open Applications, Close Applications, Run any media file, All this can be performed on the voice commands of the end user without internet connectivity. Operations such as Browsing or Searching for any topic, using Applications that need internet connections example “Send message to ABC, Hi I’m PQR sending message to you. Using WhatsApp” or sending an e-mail to concerned person just by voice command, order food from zomato, book a movie ticket, play songs over youtube. Listen to current news etc. All these basic operations are performed by the voice commands of the end user . Google did quietly enable offline recognition in that Search update, but there is no API or additional parameters available within the Speech Recognizer Class. The functionality is available with no additional coding, however the user’s device will need to be configured correctly for it to begin working and this is where the problem. Also, Google have restricted certain devices from using the offline recognition due to hardware constraints. Which devices this applies to is not documented, in fact, nothing is documented, so configuring the capabilities for the user has proved to be a matter of trial and error. Language is not a Barrier for SOFI to understand the user voice commands for performing respective operations. Hence SOFI uses its Intelligence to store the voice commands detected by the end user with respect to the Default Commands stored in SOFI.

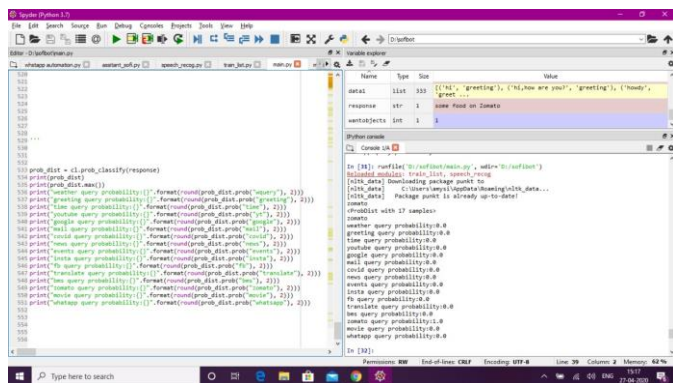


Fig-2 : Probability of intent classifier Zomato

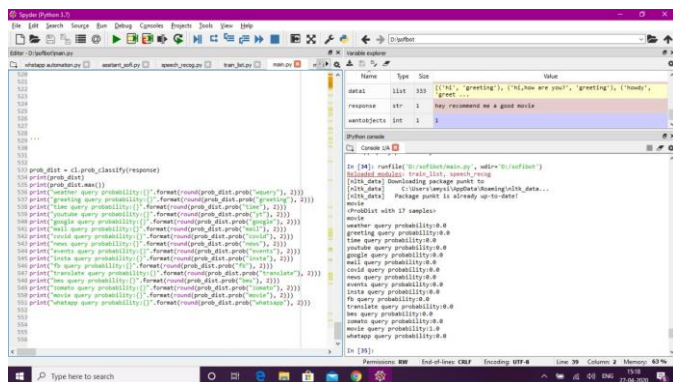


Fig-3 : Probability of intent classifier Movie

Table 1 : Probabilities of Intent Classifiers

Query	Speech Recognised	Probability of Intent Classifier
Facebook	Follow someone on facebook	1.0
Instagram	Update Story	1.0
Whatsapp	Send a message	1.0
Email	Send a mail to someone	1.0
Movie	Book a movie	1.0
Zomato	Order food	1.0

3. CONCLUSIONS

SOFI is Designed to help Native and especially for Blind persons which works on their Voice Commands. SOFI also has the capability of recognizing the voice commands without internet connection. SOFI has various functionalities of computer and web like managing various applications on just the voice commands. Contains key features like Voice Pattern Detection, Keyword Learning, etc. which helpful for end user to use various functionalities and services of the computer. Hence, SOFI is language barrier independent which actively responds to user’s voice commands faster and respond to their query.

REFERENCES

- [1] “Accurate and compact large vocabulary speech recognition on mobile devices,” in INTERSPEECH. 2013, pp. 662–665, ISCA.
- [2] “Lattice-based optimization of sequence classification criteria for neural-network acoustic modeling”.
- [3] “Speech recognition with flat direct models,” IEEE Journal of Selected Topics in Signal Processing, 2010.