

# AVAYA : SMART ASSISTANT WITH HOME AUTOMATION

Mohsin Shaikh<sup>1</sup>, Shubham Mahindrakar<sup>2</sup>, Grishma Borse<sup>3</sup>, Shraddha Aute<sup>4</sup>, Prof. Shwetkranti Taware<sup>5</sup>

<sup>1,2,3,4</sup> Undergraduate Students, Dept. of Computer Engineering, Indira College of Engineering and Management, Pune, Maharashtra, India

<sup>5</sup> Professor, Dept. of Computer Engineering, Indira College of Engineering and Management, Pune, Maharashtra, India

\*\*\*

**Abstract-** AVAYA, a voice assistant programmed using python and combined with AIML, TTS and cascades. The assistant incorporates AIML and one of the most popular and most used Text-to-Speech conversion by a leading industry-GOOGLE. Using the following libraries mention ahead, AVAYA is programmed in such a way to provide better human machine interaction with some personnel security attributes added such as Face-Recognition and Voice-Recognition.

*Index Terms:* AIML, Cascades, Python -v, SpeechRecognition, pyTorch, LBPH, opencv, Bag of words

## 1. INTRODUCTION

This is a system to intelligently speak with the users using natural language. It uses text to speech conversion api provided by Google to recognize what a person is saying and accordingly produces the outputs. The input speeches have been divided into different categories, example- "HEY!", "How are you" belongs to type of greeting speech. For this purpose json files are used. For the security or authentication purpose Face detection and recognition is used. Basically images of the user are stored in the database which are used to train the program. The images are processed and a training data model file is created that is further used to recognize a particular user. The images are converted to gray image and then the data is generated for that image in the form of LBPH (Local Binary Pattern Histogram) that basically uses a 3X3 matrix to store information.

### 1.1 Text To Speech

The first most requirement is the speech. The speech of the user is to be recognized and hence the audio is to be understood and converted to text.

There are a number of processes that are to be done to convert a speech into text. But as AVAYA is developed in python, there is no need of doing this whole of the process. Python provides a package that helps in conversion of speech to text. Python SpeechRecognition provides us everything that is needed for the job. We have functions to

recognize and convert the speech into text. It provides seven different api to covert speech to text. The speech will be recorded using the default microphone of the system and then using one of the api provided by Python Speechrecognition to convert it into text and then do the further processing according to the command given.

### 1.2 Face Detection and Recognition:

Security is one of the important factors of a system. The data should not be breached or accessed by an unauthenticated user. Face recognition is one of the widely used option for the authentication purpose. Hence a face must be detected and the recognized for a particular user. For the recognition purpose the machine needs to be trained for each user to be recognized. For the training purpose there are two different learnings- Supervised and Unsupervised Learning that are discussed further.

### 1.3 PyTorch

Pytorch is an open source library used for machine learning which is based on torch library. It has applications in computer vision and natural language processing. It is an optimized tensor library used for deep learning that uses CPU and GPU. Basically pyTorch has a class called 'Tensor' (torch.tensor) similar to numpy array to store and operate homogeneous multidimensional rectangular array of numbers.

Pytorch provides a torch.nn module to help us create and train our neural network. It is package which is used for building layers of a model.

## 2. SYSTEM MODULES

### 2.1 Authentication:

The authentication works on the principal of face detection and recognition. A user will be registered i.e sample images of the person to be registered will be stored in the database and used to train the machine so that when next time the person is authenticated the current reading images will be compared with the data that was generated during the training process and the face will be recognized.

For the training purpose LBP(Local Binary Pattern) is used. Python provides cascade classifiers that are used to do the job. We have a XML file that provides training data model for the system to detect a face. Haarcascade classifiers use LBP (Local Binary Pattern) to store the data. It stores pixel values in the form of 3X3 data matrix.

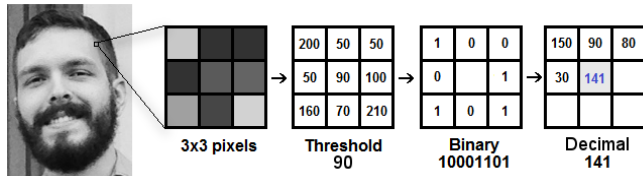


Figure 2.1.1 Pixel Matrix

LBP is a very efficient texture operator which labels the pixels of an image by thresholding the neighborhood of each pixel and considers the result as a binary number.

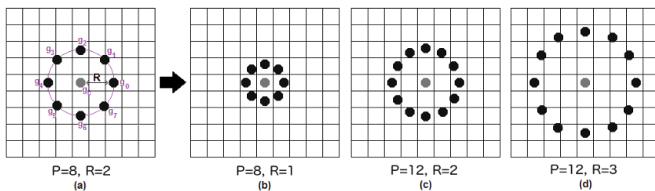


Figure 2.1.2 Local Binary Pattern

It use LBPH (Local Binary Pattern Histogram) algorithm to extract the image information and perform the matching with the input image. Using the LBP combined with histograms we can represent the face images with a simple data vector.

When for the authentication purpose the real time images are captured, such data vector/patterns are created for it and then these patterns are compared with the patterns that are present in the database file containing training model and if found so the label is returned for that image.

## 2.2 Chat Bot:

The system will be interacting the user using voice input as well as voice output. The voice input will be first recognized using an API provided by one of the world's leading organizations – GOOGLE, i.e recognize\_google(). It will recognize the voice input and convert it into text and this text will be further used to understand the input meaning and process output accordingly.

### 2.2.1 Bag of Word Model:

For the purpose of understanding of meaning of the input there are different models that can be used. "We should all believe in Humanity". The above sentence was easy for us to understand, machine cannot understand this sentence as this data is in raw form. So, we process the sentence convert the sentences in numerical format so that the machine should understand. So, for this process Bag of Words (BoW) comes into picture.

Bag of word model is a way of representing text in numerical format when modelling text with some machine learning algorithm.

How bag of word is used in our system:

Here are the patterns of inputs given to the system:

- Pattern 1: How are you?
- Pattern 2: What can you do for me?
- Pattern 3: Who are you?

So, we can't simply give these sentences to a machine learning model and ask for desired output instead we need to perform some preprocessing on the sentences before we give it to the machine learning model.

Now we first build a vocabulary or all the unique words in each pattern. Here the vocabulary consists of 9 words i.e. 'How', 'are', 'you', 'What', 'can', 'do', 'for', 'me', 'Who'.

Now we create a vector of 1's and 0's and marks the occurrences of each words with 1 and 0 respectively. Here we will have 3 vectors for 3 patterns. Refer Table 2.2.1

- Vector of Pattern 1: [ 1 1 1 0 0 0 0 0 ]
- Vector of Pattern 2: [ 0 0 1 1 1 1 1 1 0 ]
- Vector of Pattern 3: [ 0 1 1 0 0 0 0 0 1 ]

	1 How	2 are	3 you	4 what	5 can	6 do	7 for	8 me	9 who	Length of the pattern (in words)
Pattern 1	1	1	1	0	0	0	0	0	0	3
Pattern 2	0	0	1	1	1	1	1	1	0	6
Pattern 3	0	1	1	0	0	0	0	0	1	3

Table 2.2.1 Vector Pattern for Bag of Word Model

### 2.3 Home Automation:

The program will switch to home control mode if a particular command is given to the System. The home control will control smart devices using voice control. The devices that can be controlled with voice commands are

given particular unique labels to be identified by the program, so to control the device for which the command is given. We will be using arduino uno and Bluetooth module for the need. The command signals will be passed to arduino using Bluetooth connectivity.

As of now the machine is trained using labels that are returned by the program during execution. Hence when a command is given to the home control mode, it returns the tag value which is a string that is used to give signal to the arduino. The string is placed on buffer and transmitted using bluetooth module and the arduino program reads that string and based on that produces output, i.e which pin is to be set HIGH or LOW.

### 3. Machine Learning

As the system is going to recognize face and predict outputs of given speech inputs, it has to be trained to predict the outputs. A machine can be trained in two different ways, those are called as machine learnings and the types are Supervised Learning and Unsupervised Learnings.

#### 3.1 Supervised Learning:

In supervised learning, it is like training the machine with the labels provided with the training data. Differentiating data in groups, each group is provided with labels so when the machine is trained with the data it will predict the output based on these labels.

#### 3.2 Unsupervised Learning:

In unsupervised learning, the training data does not have labels. Instead the data is divided by the machine based on patterns of the data. Example we provide photos of dogs and cats, the photos will be divide into two categories based on the pattern without label and when the input is provided the machine will look for the pattern in the input and match with the pattern that is already present and predict the output that whether it is a dog or a cat.

### 4. CONCLUSION

An advanced system that can be used for home automation with face recognition and voice based interaction techniques. It can perform tasks based on the input speech of the user. Developed using machine learning and uses pre trained models for the purpose.

### ACKNOWLEDGEMENT

We would like to acknowledge the partial support of our project guide Prof. Shwetkanti Taware, serving as a Professor in the Indira College of Engineering and Management, Parandwadi, Pune. She provided proper

guidance time to time and her suggestions stood noteworthy in making this paper complete.

### REFERENCES

- [1]<https://www.irjet.net/archives/V5/i5/IRJET-V5I5395.pdf>
- [2]<https://codoid.com/the-basics-of-software-quality-attributes/>
- [3][https://www.itm-conferences.org/articles/itmconf/pdf/2020/02/itmconf\\_i\\_cacc2020\\_01002.pdf](https://www.itm-conferences.org/articles/itmconf/pdf/2020/02/itmconf_i_cacc2020_01002.pdf)
- [4][http://ece.anits.edu.in/2019-20%20BE%20Project%20REPORTS/BCM\\_1920\\_PROJECT\\_1.pdf](http://ece.anits.edu.in/2019-20%20BE%20Project%20REPORTS/BCM_1920_PROJECT_1.pdf)
- [5]<https://www.hindawi.com/journals/sp/2020/4579291/>
- [6]<https://ieevecsb.org/wp-content/uploads/sites/45/Google-assistant-controlled-home-automation.pdf>
- [7][http://smarthome.cs.iastate.edu/documents/SRS/SRS\\_3.pdf](http://smarthome.cs.iastate.edu/documents/SRS/SRS_3.pdf)
- [8]<https://t4tutorials.com/srs-documentation-of-home-automation-and-security-system-android-project/>
- [9]<http://www.iosrjen.org/Papers/Conf.19014-2019/Volume-1/14.%2058-62.pdf>
- [10]<http://www.ir.juit.ac.in:8080/jspui/bitstream/123456789/22999/1/Wireless%20Home%20Automation%20System.pdf>
- [11] <https://www.pantechsolutions.net/google-assistant-based-voice-controlled-home-automation>
- [12][http://ijariie.com/AdminUploadPdf/Home\\_Automation\\_Using\\_Google\\_Assistant\\_ijariie8241.pdf](http://ijariie.com/AdminUploadPdf/Home_Automation_Using_Google_Assistant_ijariie8241.pdf)
- [13]<https://www.analyticsvidhya.com/blog/2020/02/quick-introduction-bag-of-words-bow-tf-idf/>