Machine Learning in Virtual Personal Assistant

Sanju Abraham Varughese, Christy Baby Thomas, Jisha J Nair

Department of Computer Applications

Saintgits College Of Applied Sciences, Kottayam, India

---***---

Abstract - This Assisting users in performing their day to day tasks is the main goal of today's Virtual personal assistant applications. Now a days many Virtual Personal Assistant applications are being developed, which are capable to understand the user's daily activities, aptitudes, preferences, and goals more accurately and predicting the user's actions in advance further perform them without user's interaction. A virtual personal assistant also known as Intelligent Virtual Assistant (IVA) or Intelligent Personal Assistant (IPA) is a software agent that recognizes natural language voice commands and completes tasks for the user as per request. VPA's are basically cloud-based programs that require network connected devices or applications to work within their capabilities.

Some of the most common popular virtual personal assistants include Alexa developed by Amazon, Cortana developed by Microsoft, Siri by Apple, Bixby by Samsung and Google Assistant by Google. Virtual Personal Assistants are more flexible and have broader and more sophisticated applications and further more smarter, since they integrate AI algorithms and machine learning capabilities and make more progressive use of context information such as localization, other services, input from the user to improve their responses. In this research paper we have proposed Machine learning technique for learning mechanism of personal assistant agent and the different models used in Virtual Personal Assistant.

Key Words: Virtual Personal Assistants, Machine Learning, General Markov Model (GMM), Multi-Modal Dialogue Systems.

1.INTRODUCTION

Virtual Personal Assistants are clearing the way to the post-apps time where individuals will tend to utilize as it were a small number of predominant applications (with specific reference to messaging stages) that will coordinated numerous capacities and be engaged by computerized collaborators. Devices such as smartphones have undergone a remarkable change over the last couple of years from basic to advanced features like personal assistant, mobile banking etc. Smartphones and devices like Spoken dialogue systems would be more user-friendly if they have a capability to learn and adapt the user's behavior, interest and perform the actions without an intervention of a user. These features can be achieved with the help of Machine learning technique. Predictive Analytics commonly known as Machine learning is the study of workstation algorithms that improve routinely through experience. They use statistics to find patterns in large amounts of data which comprises a lot of things such as numbers, words, images, clicks etc. This digitally stored data can be fed into a Machine learning algorithm which in turn improve automatically through experience.

Machine learning is the technology that supports many of the services/applications that we use today like those of Netflix, YouTube, and Spotify, search engines like Bing, Google and yahoo, social-media feeds like Facebook, Instagram, Twitter, voice assistants like Siri, Alexa, Bixby. The list goes on.

There has been also many techniques that has been used to design the VPAs, based on the application and its complexity. For example, Google Assistant developed by Google uses the Deep Neural Networks (DNN), Microsoft uses the Microsoft Azure Machine Learning Studio to improve the Cortana dialogue system used in Windows based Operating Systems. Moreover. Amazon uses the automatic speech recognition also known as ASR technology for converting speech to text, and natural language understanding (NLU) to recognize the intent of the text. In general all these corporations are trying to develop the best of the core technologies/ features for their dialogue systems, such as image/video recognition, text-to-speech, dialog management and much more.

2.Related Work

We went through multiple virtual assistants and gathered all the thoughts accumulated and following is the list of the current top five related apps that are well known and appreciably used around the world.

A. GOOGLE ASSISTANT

An artificial Google Assistant, Google-designed Intelligence powered virtual personal assistant which is available on mobile and smart home devices. Google has been praised for its availability. You can get it on your phone, your speaker, your smartwatch, your laptop, your TV, or in your car. It can connect with devices or applications that are often used, such as YouTube, Netflix, and Spotify. It's smart and well integrated with android devices

B. MICROSOFT'S CORTANA

Basically, Cortana is a digital agent which is intended to learn from the commands of the user to personalize the

perception of the user as it is used later. It even can take speech inputs or email. It also offers similar features, such as management of calendars, virtual assistant conversations, queries, etc.

C. AMAZON'S ALEXA

Amazon did a really good job in the Alexa's creation. This virtual assistant has the ability to do so, to connect our entire life to a many devices. Alexa is really good. It's popular in home automation systems as it can connect to a wide number of devices, lights, televisions, electric appliances and many more. It also notifies about the weather, news, connect to the latest music available and so on as per user request.

D. APPLE'S SIRI

Siri is one of the oldest virtual assistants in the world accessible on the market. Siri can also receive voice or text inputs and respond through the same. It will provide updates to sports, news, and weather alerts, reminders and alarms, email or call people who use hands-free mode to make life simpler. For the consumers of it, only accessible on all apple devices

E. SAMSUNG'S BIXBY

Bixby is the most interesting and smart virtual assistant developed my Samsung. It provides all of the standard functionality of a virtual assistant. But one of Bixby's coolest features is that it can recognize landmarks and other sights by just taking a picture. You may also direct your camera to an item and it will show you other pricing choices. It also provides translations in different languages, such as the Google Assistant. Restaurants can also be recommended and business near you as well as manage your schedule easily.

3. THE STRUCTURE OF GENERAL MARKOV MODEL

Before the era of Deep Learning for speech recognition, Hidden Markov Model (HMM) and General Markov Model (GMM) were the two must learn technology for speech recognition. Now, there are hybrid systems which include both the HMM with Deep Learning and there are also systems which are HMM free. We have more design varieties now. However, for many generative models, HMM and GMM remains important. VPA's works on Real time environment as its offers the desired output instantaneously without any delay. As the command is inputted via the mic, the speech or command that we had already given is first processed and then it is converted to text, and from the text the keys words are extracted which in turn check with the modules which are stored in the local hard drive. If the keywords matches with any of the present modules then that particular module will be executed, else if the key word doesn't matches with any of the modules than it will notify the user to try again or can't recognize what the user wants

4. THE STRUCTURE OF MULTI-MODAL DIALOGUE SYSTEM

Spoken and graphical interfaces have their respective advantages, which means allowing an application use both,

and also allowing the users to change modality depending on the situation may be beneficial. Systems that use more than one channels or modality to communicate information are known either multimedia or multimodal systems. Multimodal systems use a higher level of abstraction from which they generate output and to which they transform the user input. Comprising some new features along with the Multi-Modal System can introduce a Next-Generation Virtual Personal Assistants.

i. **KNOWLEDGE BASE**
There are two bases of awareness. The first is an online and the second is a local knowledge base that contains all data and details dependent on each model, such as gesture modal that include facial and body data sets, speech recognition knowledge bases, dictionaries and spoken dialogue knowledge base for Automated Speech Recognition modal, video, image body data sets for Graph Model, and some user information and settings systems.

ii. **GRAPH MODEL**
The Graph Model analyses audio and video in real time and collects video frames that are captured by the camera and the input model; then sends these frames and pictures to the Graph Model and the Cloud Server applications to analyze and recover those frames and images.

iii. **GESTURE MODEL**
The Gesture Model uses the camera and Kinect in the input model to extract the gestures of the human body and the facial expressions; then sends all data to the Gesture Model and Cloud Server applications to analyze the frames and pictures and return the results.

iv. **INTERACTION MODEL**
This model is used to provide interaction between system users and system models by collecting data from the input model and evaluating the data to be submitted for each model on the basis of its tasks, and then returning it. The outcome will be used to generate the final decision.

v. **USER MODEL**
This model has all the personal details about the users who use the system and these personal details include information such as the names and ages of different users, their interests, skills and knowledge, ambitions and plans, their preferences and their dislikes or even the data about their attitudes and experiences with the system. All these information is contained in the Knowledge Base.

vi. **INPUT MODEL**
This model will coordinate the working of all input devices used by the system to gather various data such as microphone, camera and even from Kinect. In addition, this model consist of intelligent algorithms to organize input

information before passing the data to the Interaction model.

vii. OUTPUT MODEL

In addition, this model consist of intelligent algorithms to organize input information before passing the data to the Interaction model.

5. CONCLUSIONS

Machine Learning plays a major role in development of Virtual Personal Assistant. The paths of this study regarding VPAs is intended to reveal an overview on how and to what extent these devices might be used in human-computer interaction and learning. In this connection, the working systems of the VPAs namely Apple's-Siri, Google, Microsoft Cortana, Samsung's Bixby are revised within the context of Machine Learning. This VPA system used voice, graphics, video, gestures and other forms of communication for both input and output channels. In addition, the VPAs system will be used to improve contact between users and computers. The technologies that power virtual assistants require massive amounts of data, which feeds machine learning, natural language processing and speech recognition platforms.It can also be a satisfactory approach that can be used for applications such as customer response, service agent training or education, transaction facilitation, online shopping, travel information, guidance, tutoring system, ticket booking, remote banking, travel booking, information search, stock purchases, taxi bookings, route planning, etc.

REFERENCES

[1].S. Arora, K. Batra, and S. Singh. Dialogue System – A Brief Review. Punjab Technical University.

[2].R. Mead. 2017.Semio –Developing a Cloud-based Platform for Multimodal Conversational AI in SocialRobotics. 2017 IEEE International Conference on Consumer Electronics (ICCE).

[3].K. Noda, H. Arie, Y. Suga, and T. Ogata. 2014 – Multimodal integration Learning of robot behavior using deep neural networks. Elsevier: Robotics and Autonomous Systems.

[4]. M Mc Tear 2016– The Dawn of the Conversational Interface. Springer International Publishing Switzerland 2016

[5]. Y. Nung, A. Celikyilmaz – .Deep Learning for Dialogue Systems. Deep Dialogue.

[6]. Kepuska, V., & Bohouta, G. (2018) – Next-generation of virtual personal assistants (Microsoft Cortana, Apple Siri, Amazon Alexa and Google Home). 2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC).

[7]. Goksel Canbek, N., & Mutlu, M. E. (2016) – On the track of Artificial Intelligence: Learning with Intelligent Personal Assistants. Journal of Human Sciences, 13(1), 592–601.

[8]. Mane, P., Sonone, S., Gaikwad, N., & Ramteke, J. (2017) – Smart personal assistant using machine learning. 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS)

[9]. Neeta Takawale, Shibani Kulkarni, Deepashree Mehendale, Reshma Masurekar, Shweta Agarwal – A Study On Virtual Assistant in Artificial Intelligence", International Journal of Innovative Research in Computer and Communication Engineering, vol 5, issue 9, September 2017.