ML Based Virtual Personal Assistant

Anshu Kumar Singh[1] | Dr.Bhoomi Gupta[2]

[1]B.Tech Scholar, Department of IT, Maharaja Agrasen Institute of Technology, Delhi, India
[2]Assistant Professor, Department of IT, Maharaja Agrasen Institute of Technology, Delhi, India

ABSTRACT

The menial helper is shrewd programming installed in Smartphones and other associated gadgets that demonstrations like an individual right hand by helping you moderate your numerous assignments so you can focus on the more significant things. Savvy help is the need in the journey for an innovatively ahead society. With quite a rationale, our Virtual Assistant framework is proposed. The essential thought is to guarantee the security of individual data of clients and offer exact help for different day by day parts of living through a basic yet incredible framework. Such a framework utilizes Deep Learning Algorithms to accumulate results and fundamental chatterbot frameworks to go about as an ideal ally for humans. The assistant is prepared to build up its own awareness of the content, and you can show it how to speak with individuals. On the other hand, you can educate the assistant through film exchange or play contents. Be that as it may, a human-to-human discussion is a favored method to make the most ideal profound learning assistant. Keep in mind, the more information you have, the better the viability of AI will be.

KEYWORDS: Machine Learning ,Seq2Seq Model

I. INTRODUCTION

What can the Virtual individual right hand accomplish for you? The rundown extends as the innovation pushes ahead. The brilliant colleague can find data on the web, arrange your schedule and set arrangements, show an assortment of alarms and updates, enact applications while you drive, and interface with other savvy gadgets, for example, IoT hardware.

Before, remote helpers were sponsored by basic projects, equipped for fundamental voice acknowledgment. These days, these partners run on cutting edge AI and common language handling advances. These empower the partner to comprehend common discourse and answer simply like a person.

There are a few brilliant partners available, each dependent on an alternate working framework and made for a particular gadget family. The most recent right hand to join the market is Samsung's AI-controlled Bixby programming, which is unmistakably appropriate for Samsung gadgets.

Menial helpers have been included late years to different keen gadgets at home, at the workplace, and even in vehicles, with the most unmistakable being the brilliant speaker. This is a speaker for the home which runs the brilliant colleague programming and can be enacted vocally by saying its name. Samsung's Bixby can be enacted vocally and by messaging, much the same as you contact a companion.

Before, any individual who needed to speak with a PC program needed to utilize a rundown of predefined orders in a language that was more PC like than characteristic. Today, remote helpers are equipped for learning our language and even adjust to singular clients.

An extraordinary individual associate will have the option to utilize other cell phone applications to spare time and make life simpler for its client. For instance, by utilizing Bixby's AI, the Galaxy S9–Samsung's most recent leader cell phone can take better pictures with its serious camera. Bixby can likewise interpret any content caught by the Galaxy's camera and show data on areas and organizations appeared in the image. Bixby can likewise run increased reality (AR) applications. These capacities are concentrated under the Bixby Vision stage Samsung's most recent AI advancement. This capacity lets Bixby recognize cafés, attractions, and organizations utilizing AR, distinguish undertakings and waypoints, perform online quests and easily associate the actual world and the web through the camera.

Our every day reality gets effectively overburdened, and the advances that will make better arrangements are those which will take a functioning part in our lives, and will mold our future way of life.

II. Literature Review

The Internet has changed the way we live and work tremendously. Online jobs are experiencing a boom nowadays as Internet coverage and smartphone usage keeps increasing globally, by the second. Technology has become so powerful that even our concept of work has changed. More and more professionals are opting for a flexible, work from home job that allows them to seamlessly connect their work and personal lives, a feat difficult to attain with an eight to five office job.

Radio Rex was the first voice activated toy released in 1911. It was a dog that would come out of its house when its name is called.

In 1952 Bell Labs presented "Audrey", the Automatic Digit Recognition machine. It occupied a six- foot-high relay rack, consumed substantial power, had streams of cables and exhibited the myriad maintenance problems associated with complex vacuum-tube circuitry. It could recognize the fundamental units of speech, phonemes. It was limited to accurate recognition of digits spoken by designated talkers. It could therefore be used for voice dialing, but in most cases push-button dialing was cheaper and faster, rather than speaking the consecutive digits.

Another early tool which was enabled to perform digital speech recognition was the IBM Shoebox voice-activated calculator, presented to the general public during the 1962 Seattle World's Fair after its initial market launch in 1961. This early computer, developed almost 20 years before the introduction of the first IBM Personal Computer in 1981, was able to recognize 16 spoken words and the digits 0 to 9.

The first natural language processing computer program or the chatbot ELIZA was developed by MIT professor Joseph Weizenbaum in the 1960s. It was created to "demonstrate that the communication between man and machine was superficial". ELIZA used pattern matching and substitution methodology into scripted responses to simulate conversation, which gave an illusion of understanding on the part of the program.

The first modern digital virtual assistant installed on a smartphone was Siri, which was introduced as a feature of the iPhone 4S on 4 October 2011. Apple Inc. developed Siri following the 2010 acquisition of Siri Inc., a spin-off of SRI International, which is a research institute financed by DARPA and the United States Department of Defense. Its aim was to aid in tasks such as sending a text message, making phone calls, checking the weather or setting up an alarm. Over time, it has developed to provide restaurant recommendations, search the internet, and provide driving directions.

Table of Survey

S.NO	Event	Year	Purpose	Subject/Finding
1	K. H. Davis, R. Biddulph, and S. Balashek develop voiceRecognizer for digits	1952	AutomaticRecognition of Spoken Digits	The recognizer discussed will automatically recognize telephone-quality digits spoken at normal speech rates by a single individual, with an accuracy varying between 97 and 99 percent.
2	Applied Voice Input Output Society (AVIOS) is founded	1981	AVIOS provides a forum for promoting practical applications of advanced speech technology, such as speech recognition, text-to-speech synthesis.	Inform and educate developers and designers of speech applications, researchers, students, industry suppliers, companies using speech technology in their business, and individuals for whom the technology is of critical importance.
3	Dr. James Baker & Dr. Janet M. Baker Found Dragon Naturally	1982	Dragon Naturally Speaking uses a minimal user interface. As an example, dictated words appear in a floating tooltip as they are spoken (though there is an option to suppress this display to increase speed)	speech recognition software package developed by Dragon Systems of Newton, Massachusetts, which was acquired first by Lernout & Hauspie Speech Products and later by Nuance Communications.
4	Mike Phillips and Bill O'Farrell Found SpeechWorks	1994	SpeechWorks was a company founded in Boston in 1994 by speech recognition pioneer Mike Phillips and Bill O'Farrell. The	The company's main focus was bringing speech recognition solutions to phone systems. Carriers and voice portals were able to use these speech-activated

			Boston-based company developed and supported speech-related computer software.	services to direct consumer calls
5	Microsoft Speech API	1995	The Speech Application Programming Interface or SAPI is an API developed by Microsoft to allow the use of speech recognition and speech synthesis within Windows applications.	The Speech API can be viewed as an interface or piece of middleware which sits between applications and speech engines (recognition and synthesis).
6	Tellme Networks Is founded	1999	Tellme established an information number which provided time-of-day announcements, weather forecasts, brief news and sports summaries	Tellme announced a service that delivered content to telephones—a concept called voice portals.
7	James Siminoff and William Price found PhoneTag	2004	PhoneTag is a uses advanced technology to convert voicemail to text and deliver it via e-mail and or text messages.	PhoneTag is a uses advanced technology to convert voicemail to text and deliver it via e-mail and or text messages.
8	Microsoft Speech Server 2004 R2	2005	Microsoft Launches Microsoft Speech Server 2004 R2 With Expanded Language Support and New Deployment Options	Microsoft Speech Server 2004 R2 is the next step in reducing the cost and complexity of building and deploying enterprise speech solutions for customers
9	Siri	2010	Siri is published as a mobile app on iOS	Siri is a virtual personal assistant on your phone. You ask Siri in

				your own voice, and it helps you get things done when you're on the go.
10	Google Announces Voice Search	2011	Voice Search for Desktop	Voice Search for Desktop
11	Amazon Acquires Yap	2011	Yap Speech Cloud was a multimodal speech recognition system	acquired by Amazon in September 2011 to help develop products such as Alexa Voice Service, Echo, and Fire TV.
12	Microsoft Announces Cortana	2014	Cortana is a virtual assistant developed by Microsoft, which uses the Bing search engine to perform tasks such as setting reminders and answering questions for the user.	It has been launched as a key ingredient of Microsoft's planned "makeover" of the future operating systems for Windows Phone and Windows
13	Amazon Echo launches	2015	Amazon updated its cloud-enabled interactive	Amazon updated its cloud-enabled interactive speaker with the ability

III. EXISTING SYSTEM

The overall system design consists of followingphases:

(a) Data collection in the form of speech.
(b) Voice analysis and conversion to text
(c) Data storage and processing
(d) Generating speech from the processed text output , the information is gathered as discourse and put away as a contribution for the following stage for preparing. In second stage, the information voice is constantly prepared and changed over to message utilizing STT. In next stage the changed over content is investigated and prepared utilizing Python Script to recognize the reaction to be taken against the order. At long last once the reaction is recognized, yield is produced from basic content to discourse change utilizing TTS.
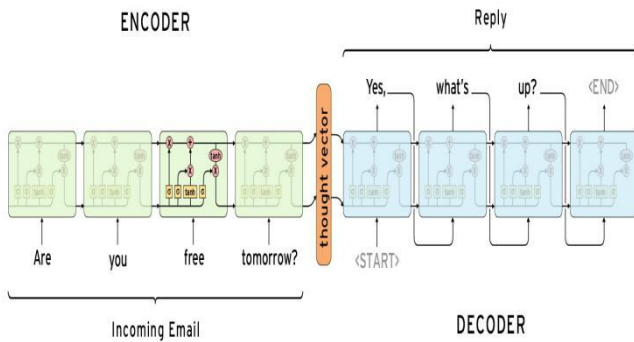
IV. PROPOSED SYSTEM
Sequence-to-Sequence Model

The point of the model is to parse the common language input x and foresee the right program y, spoken to in a consistent structure. The model takes in inputs x as a grouping of words $x = x_0x_1 \cdot x|x|$ and predicts yields y by tokens left to right $y = y_0y_1 \cdot y|y|$.
The models parts the learning into an Encoderwhich encodes x into a vector portrayal Ex by applying changes to the information grouping word embeddings we(x) and a Decoder which figures out how to create the yield tokens in the

grouping yj by utilizing the encoded data Ex and the data Dy<j from the recently decoded tokens y0y1 · yj−1.

Arrangement to-grouping learning (Seq2Seq) is tied in with preparing models to change over successions from one space (for example sentences in English) to groupings in another space (for example similar sentences meant French).

In the overall case, input arrangements and yield successions have various lengths (for example machine interpretation) and the whole information arrangement is needed to begin foreseeing the objective. This requires a further developed arrangement, which is the thing that individuals usually allude to while referencing "grouping to succession models" with no further setting. Here's the manner by which it works:

A RNN layer (or stack thereof) goes about as "encoder": it measures the information succession and returns its own inside state. Note that we dispose of the yields of the encoder RNN, just recuperating the state. This state will fill in as the "unique situation", or "molding", of the decoder in the following stage.

Another RNN layer (or stack thereof) goes about as "decoder": it is prepared to foresee the following characters of the objective succession, given past characters of the objective grouping. In particular, it is prepared to transform the objective groupings into similar successions however balance by one timestep later on, a preparation cycle called "instructor compelling" in this unique situation. Critically, the encoder utilizes as introductory express the state vectors from the encoder, which is the way the decoder gets data about what it should produce. Viably, the decoder figures out how to create targets[t+1...] given targets[...t], molded on the info succession.

TRAINING THE SEQ2SEQ MODEL

Since we made Seq2Seq model, weneedto follow the preparation process.We will test the associate at various focuses insider savvy through an info string.

At first, a large portion of your reactions will be clear as the VPA will just yield the cushioning and EOS tokens. In the end, your VPA will begin replying with little yield strings, for example, LOL, which are utilized often.

Gradually, the VPA will start building up its reactions and think of longer and more complete sentences. We will find that the appropriate responses will have a superior structure and punctuation over the long run.

V. CONCLUSIONS AND FUTURE SCOPE

Voice Controlled Personal Assistant System will utilize the Natural language handling and can be incorporated with man-made consciousness procedures to accomplish a brilliant collaborator that can control IoT applications and even fathom client inquiries utilizing web look. It very well may be intended to limit the human endeavors to collaborate with numerous different subsystems, which would some way or another must be performed physically. By accomplishing this, the framework will make human life agreeable. All the more explicitly, this framework is intended to interface with different subsystems keenly and control these gadgets, this incorporates IoT gadgets or getting news from Internet, giving other data, getting customized information spared already on the framework, and so on The android application should let the client add information, for example, schedule sections, set caution, or even updates. The product will encourage simple entry to different gadgets and stages.

The framework will have the accompanying stages: Data assortment as voice; Voice examination and change to message; Data stockpiling and handling; producing discourse from the prepared content yield. The information produced at each stage can additionally be utilized to discover designs and recommend client later. This can be a significant base for man-made consciousness machines that learns and get clients. In this manner, based on writing overview and by examining the current framework, we have reached a resolution that the proposed framework won't juststraightforwardness

to cooperate with different frameworks and modules yet additionally keeps us coordinated.

There is still a great deal of ground to be concealed in the realm of robotization however the abilities of the gadget can assist with building another age of voice controlled gadgets and bring another continuing change in the field of mechanization. This paper can likewise go about as a model for some serious applications.

REFERENCES

[1] https://www.technoarete.org/common_abstract/pdf/IJERCSE/v6/i7/Ext_23897.p

[2] https://web.stanford.edu/class/archive/cs/cs224n/cs224n.1174/reports/2748325.pdf

[3] https://www.squarefishinc.com/history-of-virtual-assistants/

[4] https://towardsdatascience.com/how-to-implement-seq2seq-lstm-model-in-keras-shortcutnlp-6f355f3e5639 https://nlp.stanford.edu/~johnhew/public/14-seq2seq.pdf

[5]https://en.wikipedia.org/wiki/Virtual_assistant#:~:text=The%201990s%20digital%20speech%20recognition,as%20we%20know%20them%20today.