



Academic Year	Module	Group	Assessment Type
A19	Complex System	L6CG3	Report

Report on Virtual assistant

Student ID : 2052267

Student Name : Manish Darji

Section : L6CG5

Module Leader : Rupak Koirala

Module Lecturer : Rohit Parajuli

Submitted on : 3rd January 2022

Contents

1. Introduction..... 1

 1.1. Why Virtual assistant? 1

 1.2. How does Virtual assistant solve the problem that mentioned above? 1

2. Aim and objectives 2

3. Literature Review 2

4. Analysis of Findings 6

5. Conclusion 7

References..... 8

Figure 1:Architecture for automatic speech recognition 3

Figure 2:Extraction Process of the Sources 4

Figure 3: Overall architecture of the proposed System 5

Figure 4:Accuracy and Performance comparison from different model of deep learning 6

1. Introduction

Virtual assistant is a man-made intelligence application software that interprets the client's voice command using standard language processing and carries out the specific task that the customer has specified in normal language. In simple terms, the associate deciphers spoken language, interacts with clients in a human-like manner, and performs a variety of tasks for their benefit.

1.1. Why Virtual assistant?

Because Virtual assistant helps to solve the following Problem that are mentioned below: -

1. Disabilities People: People who've lost their finger and hands due to a variety of reasons and must rely on their voices to do specialized tasks, such as typing (taking note, sending message).
2. Poor health: People who have a weak physical body and well-being, necessitating the assistance of devices or a partner to complete their task from the voice, as stated.
3. Repetitive Task: Nowadays, individuals have a lot of work in the government and company, and some of it is repetitive, such as writing letters to employees or people (thank you letter, important letter, etc.).
4. Busy schedule: There are many schedules in company (meetings, presentations, and so on), thus without effective management, the task or work would be ruined.
5. Time loss: Without time management in daily life, you will lose things that are vital to you, such as (meeting, events, others, etc.)

1.2. How does Virtual assistant solve the problem that mentioned above?

For the main issue involving persons with impairments and poor health who are unable to use their hands, physical bodies, and are in poor health, such as (sick patients, elderly people) who are unable to walk freely and lack the endurance to do specific duties. For example, a virtual assistant will aid them in performing chores utilizing the user's voice command (taking notes by voice command, sending message to their friends, calling, turn on and off different IOT device that related to technology etc.)

Secondary issues such as repetitive tasks, busy schedules, and time loss can be addressed by using a virtual assistant to manage time for meetings, presentations, and events. By giving voice commands to the virtual assistant, the program will manage the time and schedule according to the user's preferences. As a result, with the assistance of a virtual assistant, time and schedule will be controlled.

2. Aim and objectives

The virtual assistant's aim is to serve persons with impairments by doing specific tasks and to make work easier by performing repetitive tasks. To create an AI virtual assistant that is user-friendly.

Objective:

1. To comprehend natural language
2. To give a service that is distinct from the spoken command, for example (sending the message, doing certain task etc.)
3. To give a user-friendly interface.
4. To execute various data input, scheduling management, information search, and other relevant tasks.
5. Using a variety of algorithms to optimize the virtual assistant system
6. To utilize a secure database for storing user and system information.
7. By using the microphone and speaker to recognize the user's voice.
8. To use deep learning for speech recognition.
9. A development and testing toolkit, such as a cloud for system testing.
10. To use dataset, to train the model

3. Literature Review

As we all know, virtual assistants require a large amount of data to teach them to understand the meaning of words. It will go through many stages to comprehend human natural language. Automatic voice recognition is required to convert spoken words into text to process human speech. Because spoken signals have a high survivorship, this is a tough task. To overcome this, we employ deep learning algorithms such as deep neural networks (DNN) and deep belief networks (DBN) for automatic speech recognition, which have high accuracy and recognition rates.

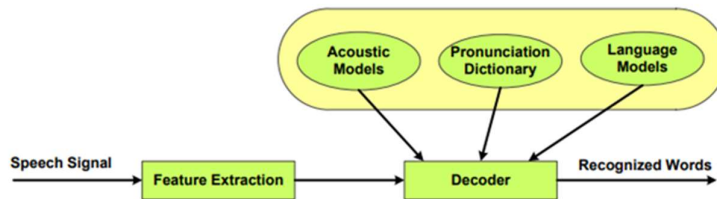


Figure 1:Architecture for automatic speech recognition

The process of speech is depicted in the diagram above, where signals pass through many stages such as feature extraction, where the dimensionality reduction approach is used to break down a vast collection of raw data into smaller, more manageable chunks for processing. It will next go to the decoder, which will transform speech to text by identifying the word. (Zhang, 2020)

1. Project Nethra - An Intelligent Assistant for the Visually Disabled to Interact with Internet Services

An intelligence virtual assistant proposed by the Nethra where (IVAs) can listens to and recognizes what the user says, and replies to the user's demands in a courteous, effective manner via voice in a conversational manner. While developing IVAs it had used four model for the processing of the speech search as voice recognition module, conversational agent, the natural language processing module for the processing natural language, the content extraction module for the extraction of the voice into raw data. In the Natural language processing (NLP) it had used the Named entity recognition to identify the People, organizations, location names, stock ticker symbols, and specific abbreviations which are the examples of named text characteristics. After that it had used Conversational Handling and Text-To-Speech modules for the to generates appropriate answers to user inputs. An input-response sequence can be thought of as a succession of volleys in a dialogue. Conversational agents have been implemented using a variety of technologies, beginning with AIML, which is an XML-based language. After that it will process data to the text-to-speech model to response the user through voice. With this architecture it had provide 80% accuracy on the speech recognition. (A.M. Weeratunga, S.A.U. Jayawardana, Hasindu P.M.A.K, W.P.M. Prashan and S. Thelijjagoda, 2015)

2. The paper distributed by Muangkammuen, et al. shows the execution of an Automated Thai FAQs (Frequently Asked Questions) Chabot utilizing the Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) design for the text grouping. The model proposed in this paper incorporates 3 layers for example the inserting layer, which is the NLP module, the RNN-LSTM layer, and the Dense (Output) layer. The NLP module maps the jargon from the corpus of printed content into the fixed-length vector portrayal. The subsequent layer, LSTM empowers RNN to handle the successive information. The last result layer utilizes the

SoftMax initiation work for use in multiclass order techniques, to get the odds of each info that has a place with a mark. The figure underneath shows the neural organization layers in this mode: Neural Network Layers. For the datasets for this model, 2,636 sets of Thai inquiries and arrangements have been arranged into eighty classes, each marked with a whole number. 60% of the information was parted for the preparation set, 20% for the approval set, and the last 20% for the testing set. The model showed 83.9% precision with the test set of information. Due to the utilization of the SoftMax initiation work, the appropriate responses were given when the result likelihood is more than the edge worth to create exact arrangements. Subsequently, 13.64% of inquiries have been forgotten about as the right answers for the inquiries were not found. Nonetheless, on the last execution assessment of the model, 86.36% of the inquiries were handled with a precision of 93.2%. (Panitan Muangkammuen, Narong Intiruk, Kanda Runapongsa Saikaew*, 2018)

3. Voice recognition system was included in the publication "Intelligent Personal Assistant - Implementing Voice Commands allowing Speech Recognition" to develop better IVAs that employed various deep neural network learning API to interpret natural language. It uses the Blind Sources Extraction (BSE) method to extract speech from noise, which has a high accuracy of voice extraction from noise.

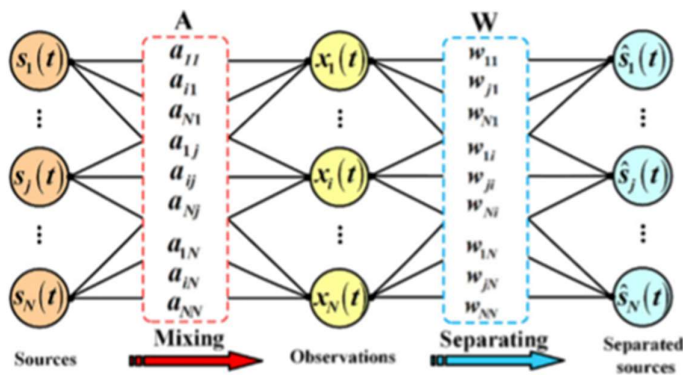


Figure 2: Extraction Process of the Sources

Sources had been mixing with A-vector, as seen in the picture, and an observation termed $X(n)$ separated the noise from the sources data into discrete sources.

It had utilized the Jurc'cek et al's technique to transform in the chunk of the word into the transitory SF (Semantic Frame) is within the terns in fragments of the words in the phrase after the sound was extracted. The activity Semantic Frame is launched when the design is discovered. The establishment is separated into two sections: the trigger component and the action portion, which operate as a locator and a change agent, respectively. Therefore, the final Semantic Frame identified in the clarified corpus is used to create it. Throughout the time spent preparing, the calculation validates the trigger-activity sets in the parser at each step. The pair that yields no activity scores, such as the one whose application yields the most basic Levenshtein distance reduction between the acquired SFs, and therefore, the specified calculation is added to

untangle the heaps and applied to the short corpus. (S[3],Rangaraj ,Kumaran N ,Siva Sharan V[2]and Dhanalakshmi R [4],, 2020)

4. Personal virtual assistants are the human future, allowing machines to communicate with humans in a human-like manner. As a result, it is critical to test the system; without testing, we cannot determine if the system is beneficial or not. Quality of Experience (QoE) is a rating provided to a system model in which the system is tested for quality and customer satisfaction with a vendor. Based on the overall testing of the system, we can determine the system's limitations and how effective our personal virtual assistant is. As a result, it provides us with feedback for the system's future improvement. (Umair Saad1 & Usama Afzal 1 & Ahmad El-Issawi2 &, 2018)

5. The NLP was used for the development of the intelligent virtual assistant in the paper "On the Track of Artificial Intelligence: Learning with Intelligent Personal Assistants," where the primary objective of the engagement is to create a representation of a text by gathering primary additions and perspectives derived from semantic information. NLP is by far the most significant component in the creation of the Virtual Assistant, as it builds the human-PC relationship for storing initial data, addressing challenges, and doing time-consuming chores requested by the user. The virtual assistant may use NLP to evaluate natural language claims, correct spelling errors, provide a syntactic framework for phrases, provide semantic relationships, and combine syntactic structure and semantic relationships for suitable responses. (Nil Goksel-Canbek2, 2018)

6. In the paper "Enabling Intelligent Environment by the Design of Emotionally Aware Virtual Assistant: A Case of Smart Campus" has proposed the Emotionally Aware Virtual assistant to make the smart camps. Where student didn't have to wait to get information from the teacher. They can simply ask the assistant to know about the information they need. While developing this system they had used the CNN and RNN-LSTM to make the system to learn about Emotion after that it is processed by NLP to the user.

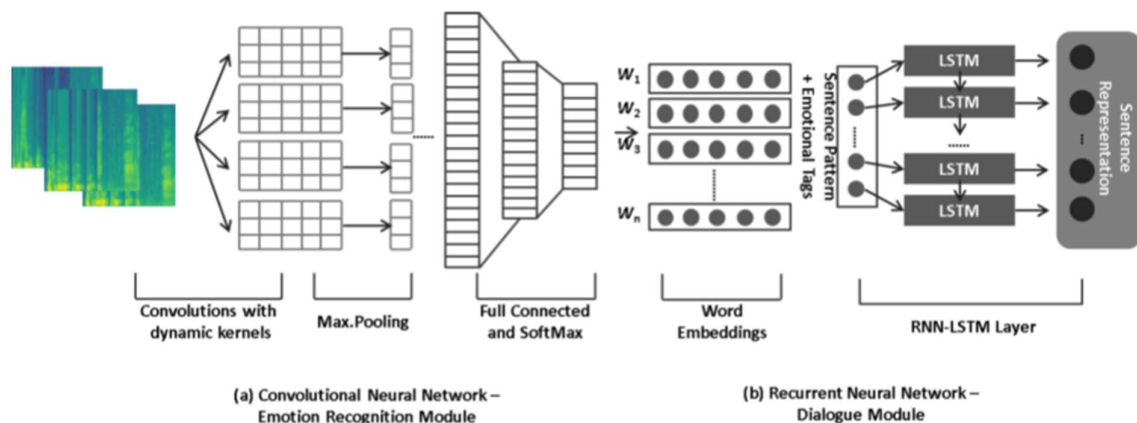


Figure 3: Overall architecture of the proposed system

As you see in the figure 3 the speech input will be transformed into spectrograms using the Fourier Transform in the first step then obtained input will process into the CNN based recognition model of emotion. Then in second part the speech-to-texts program will convert your input into words. The voice command words will subsequently be transformed into word embeddings. RNN-LSTM captures the emotional label from the emotion detection module and the word embeddings. Finally last, the natural language output will process into the voice commands to response the user in appropriate answer form. With this architecture we can get 95% accuracy and better performance in the emotionally aware Virtual assistant. (PO-SHENG CHIU¹, , JIA-WEI CHANG²,MING-CHE LEE,CHING-HUI CHEN, 2020)

4. Analysis of Findings

From the above research we had find that almost all the intelligence Virtual assistant architecture based on the deep learning from the CNN, RNN, ANN and DNN etc. some of them were modified from the original deep learning model like LSTM from RNN, R-CNN from both CNN and RNN etc. This deep learning model are trained with dataset to get better result. So here is the result that obtained from the dataset that processed into the different model of deep learning are as follows:

Models	Dataset	Performance	Accuracy
CNN	Large	98%	98%
RNN	Large	86%	89
DNN	large	95%	95%
ANN	Large	94%	95%

Figure 4:Accuracy and Performance comparison from different model of deep learning

As you see in the above figure table you can find that RNN model of deep learning have low rate of accuracy and performance according to other models because if the sequence of data is lengthy, the basic RNN becomes unable of coping with previous information, resulting in the problem of diminishing gradient during the training phase of the model that why it has low accuracy and performance. Similarly, DNN and ANN model will work on the large dataset to get better result as you see in that above table you can find that accuracy of both model is same because large dataset is processed into the model. In above table CNN have higher rate on both

performance and accuracy because this model is made for large data with more data you can get more accurate data and better performance.

Most of the system that based on the Virtual assistant has same goal to help the user in human like manner. To understand the natural language all the system had used Deep learning model to learn or to find pattern from the data to know about it. With this analysis we can find that it would be great help if this kind of system are made with Assistant, we can get lot of information from it and we can give certain task to the system to do. So, it is future of computer science technology.

5. Conclusion

According to the research, analysis, we can find that system is dependent on the good architecture and neural network, and it would be great help to the user who have physical problem, who can't see with virtual assistant he/she a trust the assistant to do certain task. To make better IVAs we need good architecture, algorithm, and dataset to process and tarin the data according to user.

References

- A.M. Weeratunga, S.A.U. Jayawardana, Hasindu P.M.A.K, W.P.M. Prashan and S. Thelijjagoda, 2015. Project Nethra - An Intelligent Assistant for the Visually Disabled to Interact with Internet Services. *2015 IEEE 10th International Conference on Industrial and Information Systems*, 18-12('1), pp. 55-59.
- Nil Goksel-Canbek², 2018. On the track of Artificial Intelligence: Learning with NLP. *Internatioanl journal of Human Sciences*, 1(1), pp. 1-5.
- Panitan Muangkammuen, Narong Intiruk, Kanda Runapongsa Saikaew*, 2018. Automated Thai-FAQ Chatbot using RNN-LSTM. *Department of Computer Engineering*, 1(1), pp. 1-4.
- PO-SHENG CHIU¹, JIA-WEI CHANG², MING-CHE LEE, CHING-HUI CHEN, 2020. Enabling Intelligent Environment by the Design of Emotionally Aware Virtual Assistant: A Case of Smart Campus. *SPECIAL SECTION ON DATA MINING FOR INTERNET OF THINGS*, 8(1).
- S[3], Rangaraj, Kumaran N, Siva Sharan V[2] and Dhanalakshmi R [4],, 2020. Intelligent Personal Assistant - Implementing Voice Commands Enabling Speech Recognition. *Department of Computer Science and Engineering*, 1(1), pp. 1-5.
- Umair Saad¹ & Usama Afzal¹ & Ahmad El-Issawi² &, 2018. A model to measure QoE for virtual personal assistant. *Multimed Tools Appl*, 1(1), pp. 1-6.
- Zhang, Y., 2020. Speech Recognition Using Deep Learning Algorithms. *AI.deeplearning*, 1(1), pp. 1-3.