

Virtual Personal Assistant (VPA) for Bengali with Machine Learning (ML) and Reinforcement Learning (RL)

Sadidul Islam
Student, MSCSE
United International University
Bangladesh
sadid@socetcon.com

Abstract—In the current technological world, virtual personal assistants (VPA) have become necessary companions, offering seamless voice-activated interaction with technologies, personalized assistance, and streamlined task management for enhanced productivity and convenience. In this paper, propose a virtual personal assistant development with speech to text, chatbot functionality, and text to speech synthesis where speech to text is implemented as Reinforcement Learning model so that it can personalize the user specific voice characteristics. By leveraging chatbot capabilities, the assistant can engage in natural language conversations, understand user intents, and provide relevant responses. The text-to-speech synthesis component ensures that the VPA can generate human-like speech output, enhancing the user experience. The developed VPA showcases the potential of combining multiple technologies to create a versatile and user-friendly virtual assistant.

I. INTRODUCTION

Technology has made remarkable strides in recent years, revolutionizing the way we live, work, and connect with each other. We can make life easier by having a personal assistant installed on a personal computer or mobile phone. All the technologies require human interaction to operate, e.g. computer, mobile, IoT devices, etc. There are situations that human can't provide proper interactions due to disabilities, lack of knowledge, and also for limited time as well. A virtual personal assistant (VPA) is built with Machine Learning technologies such as Natural Language Processing (NLP), Speech Recognition, which can help in these situations by taking a voice command from the human and completing the tasks automatically. For example, A virtual personal assistant (VPA) can be used to set a reminder, sending email, know weather information, search information in Google, update to-do list etc [1].

Numerous works for the Virtual Personal Assistant (VPA) have been done for the English language throughout recent years. Such as, Apple's Siri [9], Google's Voice Actions [10] and Google Now [11], Microsoft's Bing Voice Search [12], and Nuance's

Dragon Go! [13] and Nina [14], and many startup efforts like Speaktait [15], and many more.

Bangla is the 7th most widely spoken language in the world, 272.7 million people around the world speak in Bangla [16]. Euromonitor International reports only 18% of the total population in Bangladesh can understand and speak English [20]. In addition to that 25.34% (literacy rate 74.66% according to the Population and Housing Census 2022) of the total population is illiterate who cannot even read or write Bangla language [18] According to Unicef Bangladesh 2.8 percent of the population and 1.7 percent of children has at least one disability [17] in the country where Bengali is the primary language. Among the disabled people 11.43 (0.32 percent in whole population) percent have limitations in speaking and other 88.57 percent can at least speak, but still they are away from technologies by other disabilities.

Although, very little work related to virtual personal assistant (VPA) had been done yet [1]. We can't find any production grade virtual personal assistant (VPA) in Bengali language as of today.

In this research we propose a system with Speech to Text, Chatbot, and Text to Speech combined to develop a virtual personal assistant (VPA) that can help disabled (verbal) people to interact with technologies, save time for busy people by doing their technological tasks in no time wasted with just a voice command. This virtual personal assistant will take voice commands and based on that voice command it will do some actions like calling someone, interact with IoT devices, etc. and reply with the required information to answer the users' voice command. In other virtual personal assistant (VPA) we found that it lacks of personalizing for a specific person, but solution for this problem is more required when we consider a disabled person. Because they might not speak properly like other people, if we don't consider this problem then it won't help them at all. So, we introduce Reinforcement Learning (RL) for

Speech to Text model to personalize user specific voice characteristics.

II. REVIEW

Virtual personal assistant (VPA) has been researched and utilized by researchers over the decades. In our literature work, we focused on latest works including research, production related information which are published on reliable sources. Like, IEEE Digital Library, ACM Digital Library, Springer etc. We have found a couple of research directly related to our works and there are some partially related works. We have separated discussion of the researches in separate sections.

A. Virtual Personal Assistant

Adheetee: A Comprehensive Bangla Virtual Assistant is developed based on Natural Language Understanding (NLU) using Deep Learning (DL) models such as Recurrent Neural Network (RNN) for Bangla Language [1]. They have collect and created their own corpus then trained the model on that dataset.

In the paper "Large-Scale Personal Assistant Technology Deployment, specifically in the context of the Siri Experience" [4], the researchers provided a lot of information about development and deployment a production grade Siri, Siri is a virtual personal assistant developed by Apple. They have showed how they have implemented architecture, deployment architecture and how they have improved performance Natural Language Processing (NLP) and Natural Language Understanding (NLU) capabilities.

B. Speech Recognition

The paper State of Art Research in Bengali Speech Recognition [2] presents a comprehensive methodology for advancing the field of Bengali speech recognition. They have pointed out about challenges specific to Bengali speech recognition. The researchers proposed a hybrid approach of that combines acoustic modeling, language modeling, and lexical modeling using Convolutional Neural Network (CNN), Recurrent Neural Network (RNN) to improve accuracy and performances.

C. Chatbot

The paper "Doly: Bengali Chatbot for Bengali Education" [8] presents the methodology employed to develop an intelligent chatbot named Doly, designed specifically for Bengali language education. The methodology emphasizes the importance of continually updating and expanding Doly's knowledge base to keep up with evolving educational content and trends.

In the paper "Self-improving Chatbots based on Reinforcement Learning" [5] we have found how to

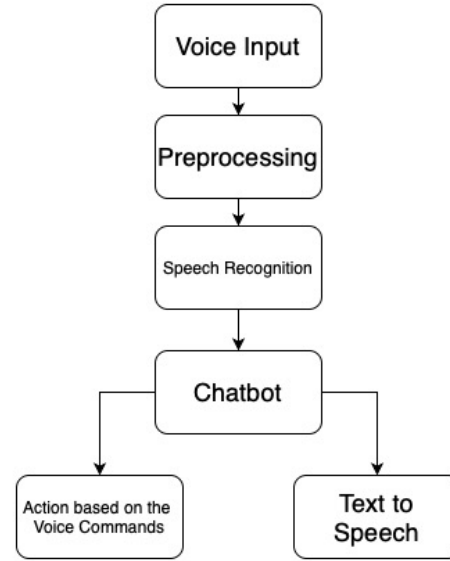


Figure 1. System Architecture of the proposed solution

create a self-improving question and answering system. They have used Deep Q-Network (DQN) agent as the policy learning system with epsilon greedy exploration. The authors propose a setup where the chatbot acts as an RL agent, observing the state of the conversation, taking actions, and receiving rewards or penalties based on user satisfaction.

III. METHODOLOGY

Our solution have three machine learning (ML) component, Speech to Text, Chatbot, and Text to Speech. Representing sounds or human voice in digital media involves converting them into analog signals. The main challenge lies in converting these analog signals into text format, which enables machines to process and work with the data effectively and the Speech to Text solves this problem [2]. A Chatbot is a computer program or an artificial intelligence which conducts a conversation via auditory or textual methods [6]. Text-to-Speech (TTS) is a technology that converts written text into spoken words [7].

Figure 1 shows a user provide a voice input, then our system process the voice data and provide that data into Speech to Text. Speech to Text returns text data for the voice data, this data is the input of the chatbot, chatbot takes the decision, like what to do, what to reply, or what to ask, these kind of questions. If the response is to take an action based on the output of the chatbot, then it takes the action, like call someone, interaction with IoT devices etc. Also, provide the voice feedback to the user using Text to Speech. There are a plenty of solutions for other top languages, but we have developed for Bengali language.

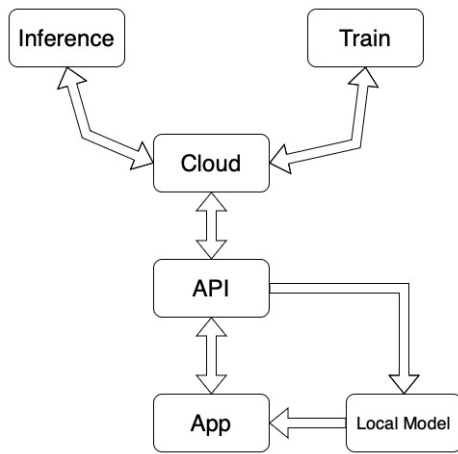


Figure 2. Deployment architecture of the system

Our deployment environment (Figure 2) is in the cloud, using API access a mobile/computer app can interact with the system. A basic model will be downloaded into the local device as well to make the communication with reduced latency. When a user provide some basic voice command the local model will be sufficient to reply or interact with that. But when a complex or unknown reply is required then it will communicate with the cloud version of the model. The models are self improvable, for example it will learn to understand a specific user in a better way by user's feedback and Reinforcement Learning (RL). Also, the local model is trained with the most commonly used commands, so that it can respond quickly for the most common words for a specific user.

REFERENCES

- [1] Syed Mohidul Islam, Most Fowziya Akther Houya, Syed Mynul Islam, Salekul Islam and Nahid Hossain, Adheetee: A Comprehensive Bangla Virtual Assistant, 2019.
- [2] S.M. Saiful Islam Badhon, Md. Habibur Rahaman, Farea Rehnuma Rupon, State of art Research in Bengali Speech Recognition, 2020.
- [3] Dipankar Gupta, Emam Hossain, Mohammad Shahadat Hossain, Karl Andersson and Sazzad Hossain, A Digital Personal Assistant using Bangla Voice Command Recognition and Face Detection, 2019.
- [4] Jerome R. Bellegarda, "Large-Scale Personal Assistant Technology Deployment: the Siri Experience, 2013.
- [5] Elena Ricciardelli, Debmalya Biswas, Self-improving Chatbots based on Reinforcement Learning, 2019.
- [6] Shafquat Hussain, Omid Ameri Sianaki, Nedat Ababneh, A Survey on Conversational Agents/Chatbots Classification and Design Techniques, 2019.
- [7] Dosik Moon, Web-Based Text-to-Speech Technologies in Foreign Language Learning: Opportunities and Challenges, 2012.
- [8] Md. Kowsher, Farhana Sharmin Tithi, M Ashrafal Alam, Mohammad Nurul Huda, Mir Md Moheuddin, Md. Golam Rosul, Doly: Bengali Chatbot for Bengali Education, 2019.
- [9] AppleSiri, <http://www.apple.com/ios/siri/>, 011.
- [10] Google Mobile, <http://www.google.com/mobile/voice-actions>, 2008.
- [11] Google Now, <http://www.google.com/landing/now>, 2012.
- [12] Microsoft Tellme, <http://www.microsoft.com/en-us/Tellme/consumers/default.aspx>, 2008.
- [13] Nuance Dragon Go!, <http://www.nuance.com/products/dragon-go-in-action/index.htm>, 2011.
- [14] Nuance Nina, <http://www.nuance.com/for-business/by-solution/customer-service-solutions/solutions-services/mobile-customer-service/nina/index.htm>, 2012.
- [15] Speaktoit Assistant, <http://www.speaktoit.com/index.htm>, 2012.
- [16] Chad Emery, "The 33 Most Spoken Languages in the World" Langoly <https://www.asme.org/engineering-topics/articles/renewable-energy/catching-the-sun> (accessed July 14, 2023).
- [17] Moyukh Mahtab, Faria Selim, "UNICEF concerned that more than half of children with disabilities in Bangladesh do not go to school" unicef Bangladesh <https://www.unicef.org/bangladesh/en/press-releases/unicef-concerned-more-half-children-disabilities-bangladesh-do-not-go-school> (accessed July 14, 2023).
- [18] Mir Mohammad Jasim, "Bangladesh's slow march towards 100% literacy" The Business Standard <https://www.tbsnews.net/bangladesh/education/bangladeshs-slow-march-towards-100-literacy-492058> (accessed July 14, 2023).
- [19] GovtBangladesh, "Report on National Survey on Persons with Disabilities (NSPD) 2021 (December 2022) [EN/BN]" reliefweb <https://reliefweb.int/report/bangladesh/report-national-survey-persons-disabilities-nspd-2021-december-2022-enbn> (accessed July 14, 2023).
- [20] Pinon, Robert; Haydon, Jon (December 2010), The Benefits of the English Language for Individuals and Societies: Quantitative Indicators from Cameroon, Nigeria, Rwanda, Bangladesh and Pakistan, Euromonitor International Ltd