Luganda Speech intent classification for IoT applications

MUGUME ELVIS 19/U/0676 Bsc. electrical engineering
KASULE JOHN TREVOR 19/U/14669/PS Bsc. Computer Engineering

PROJECT BACKGROUND

Speech command classification, also known as intent classification, is a key research area in the field of conversational AI[1]. Recent advances in conversational Artificial Intelligence (AI) have resulted in conversation-based applications with a wide range of supported platforms. **Speech recognition** is the ability of machines or programs to identify or acknowledge words and phrases from spoken language and convert them to a machine-readable format[2].

This technology has supported the growth of voice-based interactions. These systems respond to voice commands, whether they are for playing music, turning on the lights, or answering questions. In order to be effective, these systems must be able to extract the intent of the user from the voice command. This is known as intent recognition. The intent is something the user wants to do. Intent classification is a sub-task of Spoken Language Understanding and it aims to extract the aim of the user in a given voice command. Voice-based intent recognition is typically performed through automatic speech recognition (ASR) and intent classification from the transcriptions in a pipeline[3].

With the rapid increase of smart devices, there has been a growing interest in the concept of the Internet of Things (IoT). The internet has changed the way people communicate with each other by creating a virtual world for both professional and social lives during the past decades[1]. The Internet of Things (IoT) can be defined as a worldwide network of interconnected physical things. IoT can be considered an expansion of the internet that will impact our everyday lives and the way we interact with things. The foundational elements for IoT are smart objects, which are everyday items with embedded intelligence and internet connectivity. A smart object could be a smart bulb that lights when you get home.[4]

Creating a network of interconnected objects is the foundation for IoT. While this is the case, the interaction between people and the internet of things is another crucial component of building a connected life. Users can interact directly with a variety of things by creating embedded interactions within IoT objects. Enabling a Speech-based interaction with these devices improves user experience[5].

Voice commands are increasingly common in consumer electronics. Home assistants like Google Home, Siri, or Alexa[6] offer a natural voice interface to digital services and home automation with an almost non-existing entry barrier or learning curve for the user[2]. Cars offer voice interfaces to control air conditioning or media playback. This is all achievable by speech intent classification.

Work has been done in relation to speech intent classification using IoT devices. However, most of the existing work in speech intent classification and speech recognition at large is only available in major languages such as English, French and German and there has been less work done with low-resource languages. Low resource languages are those that have relatively less data for training conversational AI systems[7]. Most of the native languages in Uganda and East Africa fall under this category.

This work focuses on Luganda, the official spoken language of Buganda referring to the central region of Uganda where the largest number of native speakers are found. After English which is the official language of Uganda, Luganda is the most spoken language in the country (even more so than the second official language, Kiswahili)[8]. Luganda is spoken by about 20 million people in the country and beyond[9].

PROBLEM STATEMENT

There is increasing demand for the use of voice command systems to control devices in homes and industries. Most of the work that has been done and developed is in foreign languages and not much work has been done in native languages. There is a need to build devices that can receive and interpret voice commands in native languages. The primary focus of this work has been put on Luganda which is the most spoken language in Uganda. This project aims to implement a speech intent classification for IoT applications using Luganda voice commands.

JUSTIFICATION OF THE PROBLEM

Uganda is a culturally diverse and multilingual country with about 41 native languages[10]. All the native languages in Uganda are classified as low-resource languages in the conversational AI field[11]. This is because there are not enough resources available to work with the languages in AI tasks, especially Natural Language processing[7]. There is a steady advance in speech recognition technology throughout the world and Africa and the technology has not benefited any of these languages classified as low-resource for ASR applications. Luganda is the most spoken native language in Uganda and is used for communication in a large part of Central Uganda where technology is mostly used[10]. There is a need to include Luganda in speech recognition and voice command systems due to its great coverage in Uganda. The project will build a foundation for speech recognition tasks in low-resource languages in East Africa. The

project will also provide a basis to create smart homes and smart devices that use Luganda to be controlled.

Objectives of the project

Main objective

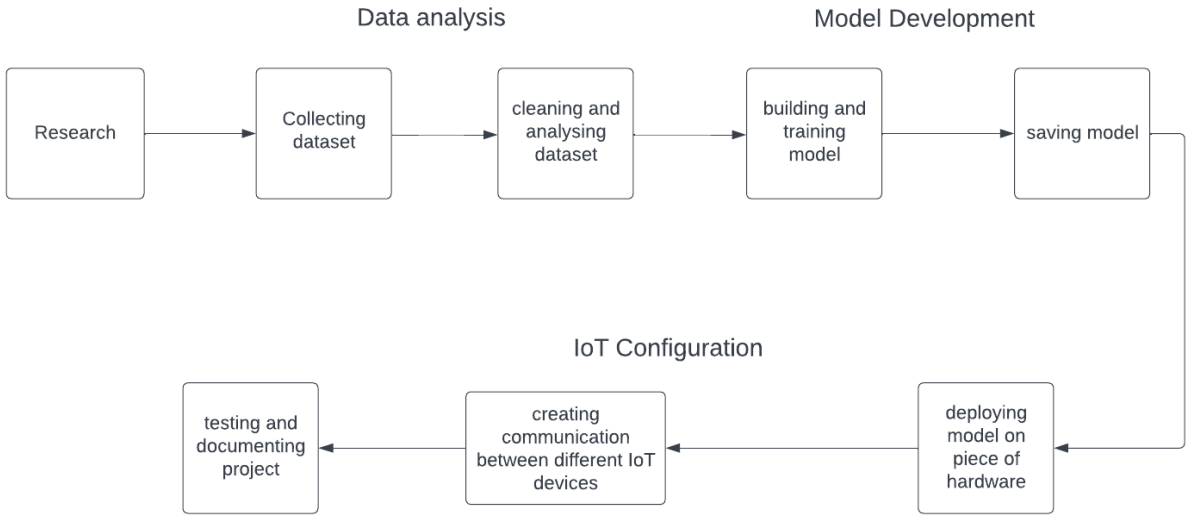
This project aims to implement a speech intent classification for IoT applications using Luganda voice commands.

Specific objectives

1. To obtain, clean, and analyze the dataset
2. To build an NLP Luganda speech intent classification model
3. To deploy the model on a hardware device with a microphone
4. To form an IoT device connection to access model on several devices.
5. To provide documentation of the project.

Methodology

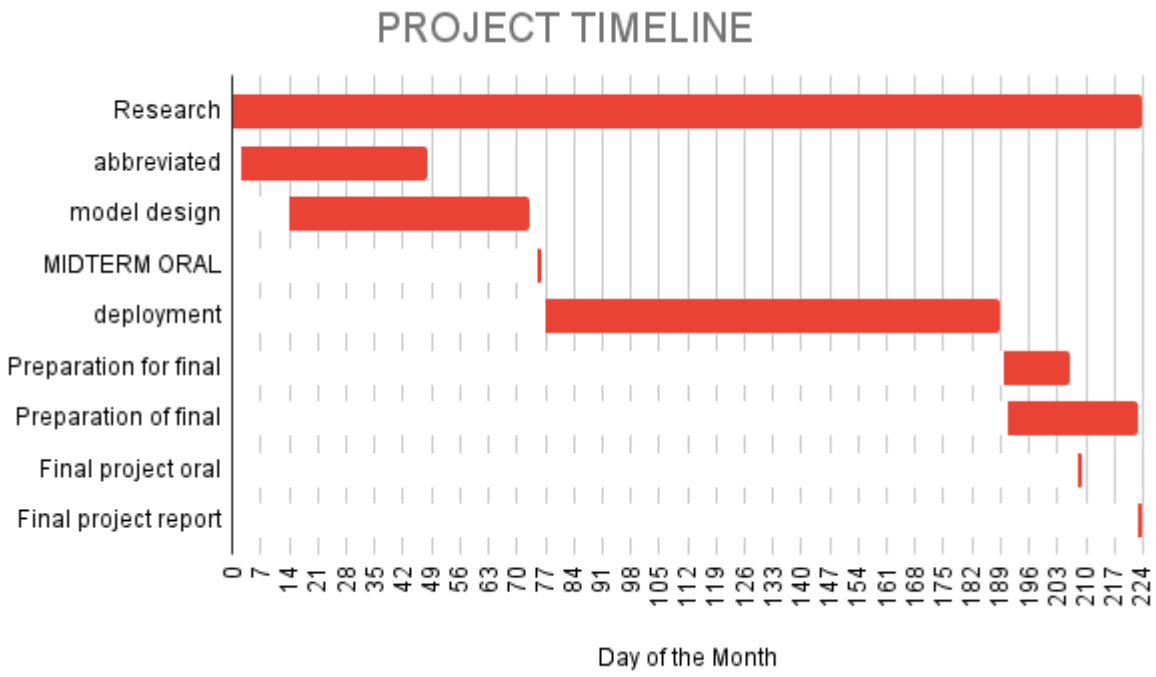
The research and literature review will be carried out throughout the project with help of journals, articles, and textbooks. The dataset will be acquired, cleaned, and preprocessed. An NLP model will be built and trained on the training dataset and evaluated on the validation dataset. The model will be placed on a hardware device with a microphone. The device will be connected to other devices with the help of an IoT connection.



Expected results

At the conclusion of the project, we hope to have a basic prototype of a network of devices that can communicate with one another and respond to a specific intent found in audio spoken in Luganda by a user into a microphone.

Timeline



Names of lecturers contacted in preparation for the proposal.

Dr. Andrew Katumba, Lecturer at Makerere university

Mr. Murindanyi Sudi, a researcher at Marconi research and innovations lab Makerere university

References

[1] T. Dinushika, L. Kavmini, P. Abeyawardhana, U. Thayasivam, and S. Jayasena, “Speech Command Classification System for Sinhala Language based on Automatic Speech Recognition,” in *2019 International Conference on Asian Language Processing (IALP)*, Nov. 2019, pp. 205–210. doi: 10.1109/IALP48816.2019.9037648.

[2] K. N., R. V., S. S. S., and D. R., “Intelligent Personal Assistant - Implementing Voice Commands enabling Speech Recognition,” in *2020 International Conference on System, Computation, Automation and Networking (ICSCAN)*, Jul. 2020, pp. 1–5. doi: 10.1109/ICSCAN49426.2020.9262279.

[3] T. Desot, F. Portet, and M. Vacher, “Towards End-to-End spoken intent recognition in smart home,” in *2019 International Conference on Speech Technology and Human-Computer Dialogue (SpeD)*, Oct. 2019, pp. 1–8. doi: 10.1109/SPED.2019.8906584.

[4] M. Mehrabani, S. Bangalore, and B. Stern, “Personalized speech recognition for Internet of Things,” in *2015 IEEE 2nd World Forum on Internet of Things (WF-IoT)*, Milan, Italy, Dec. 2015, pp. 369–374. doi: 10.1109/WF-IoT.2015.7389082.

[5] P. Wang, “Research and Design of Smart Home Speech Recognition System Based on Deep Learning,” in *2020 International Conference on Computer Vision, Image and Deep Learning (CVIDL)*, Jul. 2020, pp. 218–221. doi: 10.1109/CVIDL51233.2020.00-98.

[6] “Amazon Alexa Voice AI | Alexa Developer Official Site,” *Amazon (Alexa)*. https://developer.amazon.com/en-US/alexa.html (accessed Nov. 30, 2022).

[7] “How to Make Conversational AI Work for Low-Resource Languages.” https://keyreply.com/blog/conversational-ai-low-resource-languages/ (accessed Nov. 30, 2022).

[8] “Luganda Language: The Tongue of Buganda,” *Language Connections*. https://www.languageconnections.com/blog/luganda-the-language-of-buganda/ (accessed Nov. 30, 2022).

[9] “20 million people can speak Luganda - linguists - New Vision Official.” https://www.newvision.co.ug/category/entertainment/20-million-people-can-speak-luganda---linguis-90236 (accessed Nov. 30, 2022).

[10] “Culture in Uganda | Culture of Uganda,” *Prime Uganda Safaris & Tours*, Oct. 20, 2020. https://www.primeugandasafaris.com/culture-in-uganda (accessed Dec. 01, 2022).

[11] “Building Text and Speech Datasets for Low Resourced Languages: A Case of Languages in East Africa | OpenReview.” https://openreview.net/forum?id=SO-U99z4U-q (accessed Nov. 30, 2022).