久留米工業大学工学部情報ネットワーク工学科　2025 年度卒業研究 I

Raspberry Pi 5 を用いた話者識別対応 LLM の軽量実装と評価

概要　本研究では、Raspberry Pi 5 上で大規模言語モデル（LLM）を動作させ、音声入力による操作を実現することを目的とした。特に、使用者本人以外の音声に反応しない仕組みを構築するため、話者識別機能の実装を試みる。Kaggle から取得した公開データセットと筆者の音声を用いて CNN 型の話者識別モデルを構築した結果、学習精度は 90%以上に達したが、実際の使用では誤判定も確認された。今後は日本語対応 LLM の導入と話者識別精度の向上を目指す。
キーワード：LLM, 話者識別, Raspberry Pi 5, CNN

1．まえがき・背景

近年、人工知能技術の急速な進展に伴い、音声認識や自然言語処理を応用したインターフェースが身近な存在となっている。しかし、その多くはクラウド環境に依存しており、通信環境が不安定な場所や屋外での利用には制約がある。また、音声入力を用いるシステムは、本人以外の声や周囲の雑音にも反応してしまい、誤動作やセキュリティ上の問題を引き起こす可能性がある。

2．目的

本研究では Raspberry Pi 5 において軽量 LLM を動作させ、ローカル環境で動作可能なチャット AI の構築を行う。さらに「話者識別機能」を搭載することで、特定の使用者の声のみに反応する仕組みを確立することを目的とする。これにより、Wi-Fi 接続がなくても利用可能で、かつ誤作動の少ない安全な音声操作システムの実現を図る。

3．研究方法

研究方法は以下の通りである。

1） LLM 環境の構築

Raspberry Pi 5 上で軽量化された LLM を動作できる環境を用意する。

2） 話者識別モデルの学習用音声データ収集

学習用の音声データを用意する。

3） 話者識別モデルの構築

「利用者」と「利用者以外の話者」に分類する CNN 型の分類モデルの構築と機械学習をする。

4） 話者識別モデルの精度検証

学習済み話者識別モデルに対してテスト用の音声データを用いてモデルの精度検証を行う。

4．製作内容と評価及び研究成果

1） チャット AI の GUI アプリの作成

Raspberry Pi 5 のローカル環境で LLM を動作させる、チャット AI の GUI アプリを制作した。使用した LLM は Hugging Face にある TinyLlama を用いた。この LLM はパラメータ数が 11 億であり、他の LLM と比較して小型であるため Raspberry Pi 5 の性能であっても、実用可能な最低限の LLM であると判断し採用に至った。

アプリは python 言語で Tkinter ライブラリを用いて開発を行った。図 1 はアプリの画面とその構成を示している。上側 Text ウィジェットを入力プロンプトとし、下側には LLM の推論返答と推論時間が表示されるような設計である。

図 1　作成した GUI アプリの画面と構成

このアプリを通して、与えたプロンプトとそれに対する LLM の推論結果の 1 例から LLM の評

価を以下に示す。

入力：日本で一番高い山は何ですか？
推論：（一部略）･･･ the highest mountain in Japan is named after the Japanese language.

「富士山」という返答に期待をしていたが、実際には異なる推論結果であったことがわかる。TinyLlama の推論返答は英語での返答が主であった。日本語で与えたプロンプトに対する返答もすべて英語での返答であり、音声操作する際に非英語話者にとって不都合が生じる可能性が大きいと予測できる。また、質問の意味は多少伝わっていると考えられるが、噛み合っているとは言えない。推論精度は不十分であり改善が必要であると言える。

2)　話者識別モデルの構築と検証

システムの利用者が筆者であることを前提とし、筆者を含む 16 人分の学習用音声データを用いて「利用者」と「利用者以外の話者」に音声を分類する CNN 型のモデルの構築を試みた。筆者以外の話者の音声データは、Kaggle 上で公開されている「Speaker Recognition Audio Dataset」[1] を利用した。本データセットには 50 名分の一分長の音声データが含まれており、そのうち 15 名分、それぞれ音声データを 6 つずつ用いて学習を行った。学習に用いる音声データは、処理効率と入力長の統一を図るために数秒単位に短縮し、計算資源の少ない Raspberry Pi 上でも効率的に学習および推論を行えるようデータ処理を加えた。

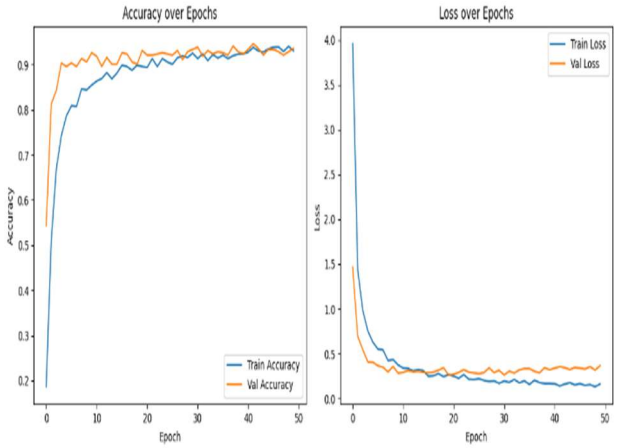

図2　CNN モデルの学習曲線（精度と損失の変化）

図 2 に学習過程における精度および損失の推移を示す。学習データと検証データの精度はともに 90%以上に到達しており、損失も低下していることが確認できる。このことから、モデルは一定の識別性能を獲得できていると考えられた。しかし、実際に筆者の音声を用いて検証を行ったところ、正しく自分として判定されず、筆者の話者として誤判定されるケースが確認された。

6．卒業研究Ⅱに向けての方針

本研究で使用した TinyLlama と構築した話者識別モデルは、実際の使用環境においては汎化性能に課題があることが明らかとなった。LLM は推論精度の高い日本語対応の蒸留モデルへの置き換えを試み、話者識別モデルはデータ量の拡充や特徴量抽出方法の改善、モデル構造の工夫などにより、特定話者の識別精度をさらに高めていく。また音声によるプロンプト入力機能の実装を目指す。

7．社会とのかかわりなど自身の研究のアピールポイント

本研究は、屋外やネットワークが不安定な環境においても利用可能な音声操作システムの実現に直結する。たとえば、車いすや福祉機器に組み込むことで、使用者が自らの声だけで安全に機器を操作でき、介護・医療分野での応用が期待できる。また、音声認証を組み合わせた制御はセキュリティ面でも重要であり、スマートデバイスや IoT 機器の不正操作防止にもつながると考える。低消費電力で動作する Raspberry Pi を活用しており、クラウドに依存しないローカル AI の新しい応用事例としてオリジナリティを持つだろう。

参考文献

[1] Kaggle, *Speaker Recognition Audio Dataset*, [online] URL https://www.kaggle.com/datasets/vjcalling/speaker-recognition-audio-dataset

[2] Raspberry Pi Foundation, *Raspberry Pi 5 Product Documentation,*[online] URL https://www.raspberrypi.com/products/raspberry-pi-5/