

1 NAIST で取り組みたい研究

1.1 はじめに

私は、貴学で「対話要約と感情フィードバックによる AI チャットボットから人への対話引き継ぎの改善」という研究に取り組みたいと考えている。図 1 に本研究の全体像を示す。

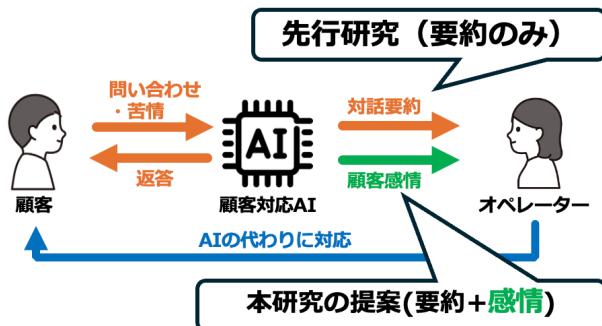


図 1: 本研究の全体像

近年、コスト削減や顧客サービスの効率化を目的に、多くの企業が窓口やコールセンターなどの顧客対応に AI チャットボットを導入している。しかし、ChatGPT などの大規模言語モデル (LLM) の発展にもかかわらず、完全に自立した AI が顧客の多様な要望に全て対応することは依然として難しい。この課題に対して、AI チャットボットによる初期の対応が難しくなると人のオペレーターに対応を引き継ぐ方式（対話引き継ぎ）がよく取られている。

対話引き継ぎに関する先行研究では、AI と顧客とのやりとりを要約してオペレーターに提示することが対話引き継ぎに有用であることが示されている [1]。顧客対応に関する先行研究では、顧客が負の感情を抱えている場合には迅速な対応が顧客満足度の向上に効果的であることが明らかとなっている [2]。また、オペレーターの認識している顧客感情と、実際の顧客感情が一致しないことが指摘されている [3]。こうした感情認識の重要性が示されているにもかかわらず、対話引き継ぎ時に顧客感情をオペレーターに表示することの効果を検証した研究はまだ行われていない。

このような背景から、対話引き継ぎにて、対話要約だけではなく、顧客感情を併せてオペレーターに認識させる仕組みが必要だと考えられる。そこで、本研究では、AI チャットボットから人への対話引き継ぎ時に「対話要約と感情フィードバック」をオペレーターに提示することで、(1) オペレーターの初動対応や対応方針がどのように変化するか、(2) 顧客満足度や全体の対応効率がどの程度改善されるかを調査する。これにより、強い負の感情を抱えた顧客にも即応しやすくなり、また、顧客感情の誤認識による不適切な顧客対応が解消され、サービス品質の向上が期待される。

1.2 提案手法

本研究では顧客対応の場面で AI チャットボットが初期対応を行い、顧客感情が悪化した場合にオペレーターに対話を引き継ぐシステムを構築する。そして、電話での顧客対応を模したロールプレイングにて同システムの有用性を確認する。

1.2.1 システム概要

本研究のシステムでは、まず AI チャットボットが顧客対応を行う。次に、システムが対話引き継ぎが必要だと判断した際にオペレーターへの対話の引き継ぎを実施する。

顧客発話の文字起こしには、高速・高精度な文字起こしが行える Azure Speech to Text API を使用する。

AI チャットボットの応答生成には、高速な GPT-4o mini を使用し、検索拡張生成 (RAG) でロールプレイングの題材に関する知識から応答を生成させる。そして、AI チャットボットの音声合成には高品質な音声合成が可能な Voice Text Web API を使用する。

音声感情認識 (SER) には、VGG-Net にランダムフォレストと多層パーセプトロンを組み合わせた軽量・高精度なアーキテクチャ [4] を使用し、顧客の発話ごとに SER を行う。

対話要約では、まず顧客の発話と AI チャットボットの応答を 1 セットに要約を生成する。次に、各ローカル要約を基に全体の要約を生成する [5]。この手法により、リアルタイムでの対話要約が可能となる。要約生成には GPT-4o mini を使用する。

対話引き継ぎのタイミングについては、システムが SER で顧客の負の感情が強くなったことを検知した場合にオペレーター側の PC にアラートを表示し、対話を引き継ぐ。その他、顧客が AI チャットボットに対して「人と交代して」などの発話を行った場合に對話を引き継ぐ（ルールベース）。

そして、対話が引き継がれる際にはオペレーター側の PC に對話要約と感情フィードバックを表示する。

1.2.2 実験に用いるロールプレイング概要

ロールプレイングは電話による問い合わせの場合を想定して、顧客役 1 人とオペレーター役 1 人で行う。図 2 に本ロールプレイングの概要を示す。



図 2: ロールプレイング概要

ロールプレイングの流れは以下のとおりである。

- (1) 顧客に聞き出してほしい情報を指示する。また、オペレーターには顧客が知りたい情報を事前に伝えておく。
- (2) 顧客が窓口に電話をかけると、チャットボットが初期対応を行う。
- (3) システムがチャットボットによる対話継続が困難と判断してアラートを発したら、オペレーターが対話を引き継ぐ。
- (4) 顧客が指示された情報を取得できたら、ロールプレイングを終了する。

上記のロールプレイングを、(1) 対話要約のみ、(2) 感情フィードバックのみ、(3) 対話要約と感情フィードバックの 3 条件で行う。

1.3 システムの評価

システムの定性的評価として、顧客役には顧客対応の評価アンケート、オペレーターには対話引き継ぎの行いやすさについてのアンケートを行い、システムが顧客満足度の向上、対話引き継ぎの円滑化にどのような影響を与えるかを分析する。

定量的評価としては、システムがアラートを発してからオペレーターが顧客対応を始めるまでにかかった時間を計測し、システムが対話引き継ぎの効率向上にどれくらい寄与したかを分析する。

1.4 研究の手順

研究は、以下の手順で実施していく予定である。

- (1) 提案システムの構築を行う。
- (2) 前述の3条件でロールプレイングを実施し、結果の分析を行う。(ここまでを、1年生の10月までに行う予定である。)
- (3) システムの改善を行い、再び手順(2)を行う。

1.5 予想される結果

本研究の予想される結果として、以下のことが挙げられる

- 感情フィードバックによって、顧客の負の感情が強い時にオペレーターが迅速な初動対応を行うようになる。
- 対話要約と感情フィードバックを基にしたオペレーターの迅速かつ適切な対応によって、対話要約のみと比べて顧客満足度が向上する。

1.6 検討事項

対話要約では、ASRの質が悪い場合に誤った要約が生成される可能性がある。Azure Speech to Text APIのASR精度が十分でない場合は、OpenAIのWhisperなど、他のモデルを使用することを検討する。

また、要約の生成にはGPT-4o miniを利用するが、要約生成がリアルタイムに行えない場合は、要約の質とのトレードオフを考慮しながら、さらに軽量・高速なLLMの使用を検討する。

2 これまでの修学内容

2.1 はじめに

これまでの修学内容として、現在取り組んでいる研究である「正常心音のみで学習したGANを用いた異常心音検出」について述べる。

心血管疾患(CVD)は心臓や血管に異常が生じる疾患の総称であり、世界的に主な死因の一つである。早期発見には専門医の聴診が有効であるが、地域によっては専門医が不足しているか、経済的理由により十分な医療が受けられない場合がある。このような背景から、CVD診断のために自動で異常心音を検出する研究が行われてきた。近年では、深層学習モデルを用いた異常心音検出の研究が進められており、多くの研究で高い検出能力が示されている。これらの研究では教師あり学習が用いられており、正常・異常の両方の心音データが

必要となる。しかし、異常心音のサンプル数は正常心音よりも少なくデータ不均衡が問題となる。

上記の課題に対し、私は正常心音のみを用いて深層学習モデルを構築する手法の開発に取り組んでおり、データ不均衡の影響を受けにくい異常心音検出の実現を目指している。

2.2 今取り組んでいる手法

本手法の異常心音検出ではまず、正常心音のみで学習した敵対的生成ネットワーク(GAN)の生成器に、メルスペクトログラムに変換した心音データを入力しデータ再構築を行う。次に、再構築データと入力データを識別器に通して得られる出力の差分(絶対誤差)を算出する。そして、この絶対誤差が閾値より小さい場合には正常心音、大きい場合には異常心音と判定する。ここで使用する閾値は、学習時にさまざまな閾値を試した中でF1スコアが最も高かったものを採用している。本手法の概要を図3に示す。

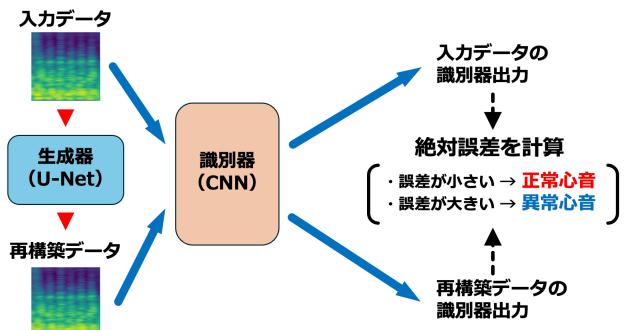


図3: 手法の概要図

現在の進捗として、PhysioNet/CinC Challenge 2016[6]で提供されている心音データセットを使用して5分割交差検証を行い、再現率: 0.67、適合率: 0.49という結果が得られている。今後はバンドパスフィルタを適用するなどのデータ前処理や学習アルゴリズムの改良を通じて、再現率、適合率共に10%を越える分類性能の向上を目指す。

2.3 おわりに

現在の私の研究は音データを利用するものであり、ここで得た経験・知識は、前章で提案したシステムの構成要素であるSERのモデル開発などで活かすことができると考えている。

参考文献

- [1] Sanae Yamashita et al., *Investigating the Effects of Dialogue Summarization on Intervention in Human-System Collaborative Dialogue*, HAI, 2023.
- [2] Oluwole Iyiola et al., *The Relationship between Complaints, Emotion, Anger, and Subsequent Behavior of Customers*, IOSR-JHSS, 2013
- [3] Anna S. Mattila et al., *The Role of Emotions in Service Encounters*, Journal of Service Research, 2002
- [4] Samson Akinpelu et al., *Lightweight Deep Learning Framework for Speech Emotion Recognition*, IEEE Access, 2023.
- [5] Khai Le-Duc et al., *Real-time Speech Summarization for Medical Conversations*, Interspeech, 2024.
- [6] Gari D. Clifford et al., *Classification of normal/abnormal heart sound recordings: The PhysioNet/Computing in Cardiology Challenge 2016*, Computers in Cardiology, 2016.