

VR空間でのChatGPTを用いたリアルタイム話題要約システム

遠藤 冴花* 三武 祐玄*

概要. 近年, VRChat や cluster といったメタバースプラットフォームの普及から, メタバース空間でのコミュニケーションも一般的になってきている. アバタを用いたり, 会話音量の距離減衰による複数の会話が並列で存在したりと, 実空間により近い形でコミュニケーションが取れるようになってきている. その一方で, 実空間で起こりうる会話上の問題も発生している. 会話グループが複数あり, 様々な話をしている中で, 自分の興味のある会話をしているグループを見つけるためには手間がかかってしまう. そこで本研究では雑談や歓談といった会話グループが複数存在する空間で, どのグループがどのような話題を話しているかを共有することで, 入りたい会話グループを見つけやすくし, メタバース上での会話をより円滑に進めることを目的とする. システムの実装はVRChat 上でを行い, 話題要約のために ChatGPT を用いて行う.

1 はじめに

メタバースの普及により, コミュニケーションを行う場は広がりを見せており, 現在ではVRChat や cluster などのメタバースプラットフォームでのコミュニケーションも一般的になってきている. アバタ同士の距離間による会話音量の減衰によって同じ空間に複数の会話が並列で存在すること, その空間への出入りが自由な点から雑談の場が自然発生する. また, 雑談を主としたイベントの開催も実際にメタバース内で盛んにおこなわれるなどコミュニケーションの場として実空間に近いものとなっている. それと同時に実空間で遭遇する会話上の問題もメタバース内の会話で起こりうる状況になっている.

本研究で扱う会話の形態は雑談や歓談といった複数人が会話に参加し, 参加している人が増加減少を繰り返すような場面を想定している. 会話グループが複数存在している空間では参加したい話題のグループを選択する必要がある. しかしメタバース空間で, ワールドに途中参加した時には既に雑談が形成されていることもあり, それぞれのグループでどのような内容を話しているのかは実際に参加してみないとわからないという事態になりえる. ここで入りたい会話グループがどれなのかを判断するために, すべての会話を聞いていくのは手間である. 本研究ではこの課題を解決するため, 会話の話題共有に着目したシステムをVRChat 上へ実装し, 「入りたい会話グループの選択」を行いやすい環境を作ること

Copyright is held by the author(s). This paper is nonrefereed and non-archival. Hence it may later appear in any journals, conferences, symposia, etc.

* 明治大学

で会話への円滑な参加にどれだけ貢献できたか評価をおこなう.

2 関連研究

文の自動要約に関して, 従来の要約においては文同士のつながりが十分に考慮されておらず, 要約文が読みにくくなる課題がある. これに対し, 要約する際に文章中に出現する名詞を「話題」として抽出し, それらがどれだけ関係しているかを計測することで文のつながりを評価する手法がある[1][2]. この評価値を基に会話がつながりやすいように話題を選択し組み合わせることで, 話題につながるのある要約文を生成している. 従来の要約文より, この手法で生成された要約文のほうが読み手にとって一貫性のある内容を提供することが実験から明らかになっている. この研究から話題要約のために会話中の「話題」を抽出するだけでなく, どのように変化していくかを考慮することでより自然で一貫性のある要約文の生成ができることが分かる.

3 話題要約システム

3.1 提案手法概要

本研究では, 「会話への入りにくさ」を解消し, コミュニケーションを円滑にすることを目的として, 後から参加する人が話題を把握しやすいように, 会話中の話題を常時要約表示するシステムを提案する. これを実現するためにはVRChat 内での会話を取得したうえで内容を要約する必要がある. また取得した要約を, のちに会話に参加する人が話題の内容を理解できるように空間内に表示しつつそれをリアルタイムで更新していかなければならない. そのため

に VRChat 内での会話取得と文字起こしを行う Whisper, 会話内容の要約のために ChatGPT を利用する。

ChatGPT は OpenAI 社が 2022 年より開始した対話型 AI サービスである。本研究では文字起こしされた会話内容の要約に, ChatGPT を使用して自然言語処理を行う API を使用する。

Whisper は ChatGPT 同様に OpenAI 社が提供している音声認識モデルである。音声を自動的に文字起こしすることが可能であり, VRChat 内での会話を文字起こしするため利用する。本研究ではリアルタイムで処理を行うため, マイクから取得した音声を直接 Whisper で文字起こしする mallorbc 氏によるオープンソースのライブラリ Whisper_mic を使用する[3]。これにより, PC のマイクから入力された音声を Whisper で文字起こしすることを実現する。このライブラリは Whisper の精度を, 動作環境に合わせて 5 段階の中から指定することができる。また, VRChat で取得した音声をマイク入力に変換するために仮想オーディオミキサーソフトである Voicemeeter Banana を使用し, 一連のシステムを Python にて行う。

3.2 提案手法概要

システムの概要図を図 1 に示す。図にある「ユーザ A,B,C」のような VRChat 内で会話している集団とは別に「聞き取りユーザ」を空間内に配置し, そちらで会話の收音を行う。聞き取りユーザは bot 的な立ち位置として, 空間内ではオブジェクトとして存在し, 会話は聞き取りユーザ付近で行うことを想定している。VRChat 上での会話音声は VoiceMeeter Banana でスピーカー出力からマイク入力に変換し Whisper で文字起こしを聞き取りユーザ側の PC でローカル処理を行う。文字起こしした内容の要約とサーバへの保存はインターネット上で実行し, VRChat からのリクエストを受けた際に話題の要約文をサーバから VRChat 側に送信することで, 話題の表示を行う。

話題の表示は, 聞き取りユーザのいる地点上空, VRChat のワールド内にいる他ユーザが視認しやすく, 話している集団に対しては視界に入らないような高さに設定する。このワールドは VRChat で利用できる HTTP でテキストを取得してくる VRCStringDownloader と Unity のテキスト表示機能を用いて Unity で開発を行った。

3.3 要約文表示までの流れ

システム実行の流れを図 2 に示す。Python 上で実行する際に Whisper_mic での文字起こし, ChatGPT での要約処理, サーバへの書き込みを並

列処理しており, それぞれ時間によって格納されている情報を処理している。本研究では Whisper_mic での文字起こしを 1 分毎に蓄積し, その後 VRChat への要約文表示の処理を行っているため, おおよそ直前 1 分間の会話文要約を表示する。

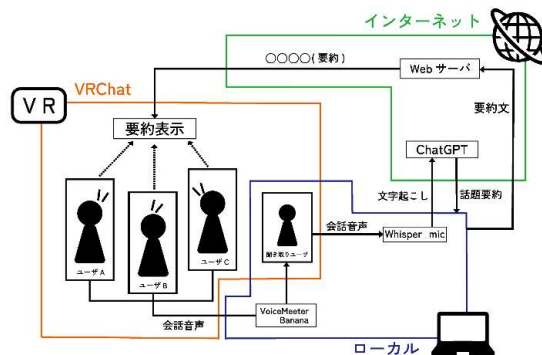


図 1. 話題要約システムの概要図

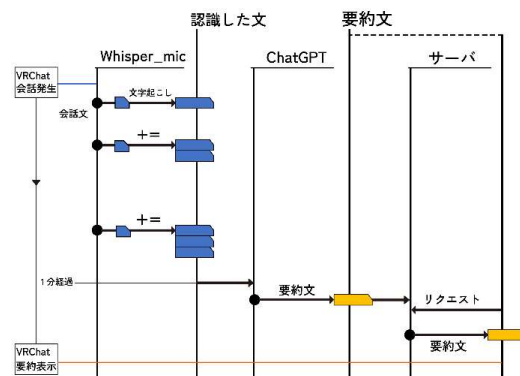


図 2. 話題要約システム実行時の時系列

4 今後の計画

今後の計画としては, 会話グループの話している内容の情報をどれだけ正確に得られているかを検証するための実験を行い, 結果を考察することを目的とする。ChatGPT から出力された要約文が実際に会話内容理解にどの程度寄与するのかを検証して行く必要があり, ChatGPT 側に送信しているプロンプトの修正を行うなどの改善が求められる可能性がある。さらに本研究で会話文蓄積に設定している“1分”の時間設定や要約文の表示方法は実験を複数回行ったうえで修正していく必要があると考えている。これら話題要約の精度に加えて, 会話内容のプライバシー問題やどのような表示方法が適切かを検討していき, 適切なシステムの実装を目指していく。

参考文献

- [1] 市丸夏樹, 飛松宏征, 日高達(2004) “話題の流れを保持する自動要約”, 情報処理学会研究報告 NL160-7,

VR 空間での ChatGPT を用いたリアルタイム話題要約システム

2004 (23), pp.43-48.

- [2] 市丸夏樹,日高達(2005) “要約文の話題の流れの最大化による自動要約.” 自然言語処理, 12 (6), pp.45-61..
- [3] mallorbc(2023),whisper_mic
https://github.com/mallorbc/whisper_mic