

どんな研究

対話システムが日常場面で人々と自然に会話するためには、そこで起きている会話の性質や、会話参加者の関係といった対話の状況を正しく理解する必要があります。本研究では、**日常場面における状況ごとの対話の分析手法と、対話の状況をマルチモーダル情報から推定する技術を開発しました。**

どこが凄い

私たちの日常会話が対話の目的や対話中の話し方に関する7つの要素で説明できることを明らかにし、**各要素の強さを会話内容から自動で測定する手法を構築しました。**また、対話場所や対話参加者の関係といった**日常会話の状況を音声・画像・言語情報を用いて認識する手法を初めて提案しました。**

めざす未来

近年、対話システムの応答の自然性は飛躍的に向上しました。ここにマルチモーダル情報処理を統合することで、状況に合わせた社会的な会話も可能になると期待されます。本技術を発展させ、**日常生活の中で私たちが対話システムと会話することが当たり前になる未来の実現をめざします。**

日常会話に参加する対話システム

■ 対話システムが私たちの日常会話に自然に加わるには、状況に合わせた会話ができる必要がある

- 対話状況に合わせた応答の例



⇒ 対話の状況を正確に理解する技術が必要

■ 日常会話の状況を理解するための2つの取り組みを紹介

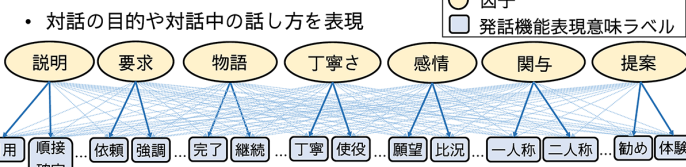
1. 日常会話の構成要素の抽出と解析 (対話内容の理解)
2. マルチモーダル情報を用いた対話状況の認識 (対話場面の理解)

日常会話の構成要素の抽出と解析

■ 日常会話コーパス[†]の対話の特徴量を因子分析し、**会話を構成する7つの因子**を抽出

- 分析の特徴量: 会話から得られた発話機能表現の意味ラベル
 - 会話の特徴を捉える有用な手がかり
 - モダリティ、主観的情報、時制などを表現

• 抽出された会話の因子



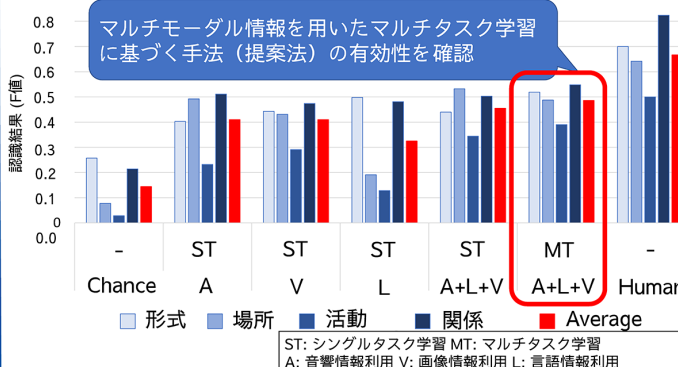
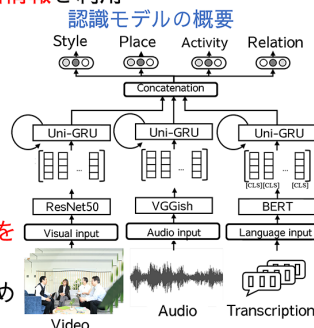
マルチモーダル情報を用いた対話状況の認識

■ 問題設定

- 短い対話シーンを観測し、対話状況を逐次分類
- 会話における**音声・画像・言語情報**を利用

■ 認識モデル

- 推定対象
 - 対話形式 (例: 会議会合、雑談)
 - 対話場所 (例: 飲食店、車内)
 - 対話に伴う活動 (例: 食事、休息)
 - 話者間の関係 (例: 家族、友人知人)
- 30秒間の**音声・画像・言語入力**を事前学習モデルで埋め込み
- 対話状況間の関係を反映するため**マルチタスク学習**



提案技術による対話処理 (デモ)

• 2つの対話状況分析技術を独立に駆動

• 対話システムの応答適応、日常会話の解析・記録への適用

関連文献

[1] Y. Chiba and R. Higashinaka, "Dialogue situation recognition for everyday conversation using multimodal information," in *Proc. INTERSPEECH*, pp. 241-245, 2021.
 [2] Y. Chiba and R. Higashinaka, "Analyzing variations of everyday Japanese conversations based on semantic labels of functional expressions," *ACM Transactions on Asian and Low-Resource Language Information Processing*, Vol. 22, No. 2, pp. 1-26, 2022.

連絡先

千葉 祐弥 (Yuya Chiba)
 協創情報研究部 実世界インタラクション研究グループ