2H4-OS-35b-1

ユーザーの態度推定に基づき適応的なインタビューを行うロボット対話システム構築への一検討

A Case Study Toward Implementing Adaptive Interview Strategy Based on User's Attitude Recognition for a Interview Robot

長澤 史記 *1 Fuminori Nagasawa　石原 卓弥 *1 Takuya Ishihara　岡田 将吾 *1 Shogo Okada　新田 克己 *1 Katsumi Nitta

*1東京工業大学 情報理工学院 情報工学系
Dept. of Computer Science, School of Computing, Tokyo Institute of Technology

The goal of this research is to develop a robot dialogue system that elicits story of users by choosing appropriate questions which user would like to answer. The key technique is the inference of users' motivation from multimodal information: gesture, posture, speaking status and prosodic status. In this preliminary research toward the goal, we develop the prototype of an interview system using humanoid robot with multimodal sensing devices.

We collecte the dialog data corpus through interview with five users and annotate the level of motivation for answering the question. We investigate the useful features to infer the motivation by student's t-test.

1. はじめに

広報や案内のような, 特定の事柄について他人に伝えるべき情報を収集する手段として, インタビューが広く用いられている. インタビューでは, 事前に収集するべき情報についての質問事項を作成して質問するだけではなく, 相手の反応によって, 話題を掘り下げる, 話題を転換するなどといった質問戦略の選択を行う必要がある. 相手に語らせることによりより, 多くの情報を引き出すため, 情報源であるインタビュー対象者の発話意欲を推定し, より意欲的な発話が得られるような話題選択を行うことも重要な技術となる. また, インタビュー結果をもとに情報を発信する際には第三者にとっても理解しやすいように具体的な情報を多く含む事が好ましい. そのためインタビュー対象者の回答内容を考慮して, より具体的な回答を得られるような戦略判断も重要な要素となる.

本研究では, インタビュー対象者であるユーザーのマルチモーダル情報から発話意欲の推定を行い, 推定結果に基づいた適切な質問戦略を選択することで, インタビュー対話によってユーザーからより深く掘り下げた内容を聞き出すインタビュアーロボットシステムの構築を目的とする.

この目的に向けて, インタビュー対話を通してユーザーのマルチモーダル情報を収集するシステムを構築し, インタビュー実験を実施してマルチモーダル情報の収集を行った. 実験で得られたマルチモーダル情報について相関分析を行い, 有用な特徴量の検討を行った.

2. 関連研究

Saito ら [Saito2015] は，高齢者との問診対話を目的とした対話システムにおいて，ユーザの発話意欲推定に関して検討した．視線の動き，フィラーなどの特徴量や，書く非言語イベントの生起タイミングが有効な特徴量であることを示したが，多くの特徴量は手動でアノテーションされており，発話意欲の自動推定は課題として残っている．本研究はユーザの態度を自動推定することを目指し，特徴量は全て自動的に抽出する．

岡田ら [岡田 16] は, 対話に参加する人間のコミュニケーション能力を評価する方法として, マルチモーダル情報を用いることが有用であることを示した. 本研究では同様の手法で測定したマルチモーダル情報を用いるが, マルチパーティ対話についての個人の能力についての事後評価ではなく, 一対一の対話において, 各個人の態度についてリアルタイムでの特徴量抽出に用いる.

千葉ら [千葉 14] は, ユーザーの対話意欲を考慮した話題選択を行うインタビュー対話システムのために, 人-人間でのインタビュー対話について分析を行った. インタビュー対話を通じてユーザーから多くの情報を引き出せるように相手の意欲の高い話題毎の対話意欲を推定するためには, 音声や顔や全身の動きなどのマルチモーダル情報を用いる事が有効であることを示した.

本研究では, 対話ロボットの質問戦略を逐次更新するために, 対話の応答毎に意欲の推定を行うほか, 骨格を考慮した姿勢に関わる非言語特徴量も考慮する. このため本研究では人-ロボット間でのインタビュー対話におけるマルチモーダル情報を取得し, 発話意欲との関連性を分析する. また, 本研究では発話意欲の尺度として, 対象が意欲的な発話を行っているかという積極性の観点からアノテーションを行う.

3. システム概要

本研究で目標とするインタビュアーロボットシステムについて説明する. 本ロボットシステムは，インタビューで話を引き出し，詳細なインタビューログデータを得て，そのデータから記事やハイライト動画を作成することを目的とする.

システムの構成を図 1 に示す. ロボットがインタビュー対象者と対面し, 一対一のインタビュー対話を行う. ロボットは質問を行い, ユーザーの回答時の状態を音声・画像・姿勢の情報から, リアルタイムで解析し, これらの情報を基に次に行う質問の選択を行う. ユーザと直接対話するロボットには pepper を使用し, 各種センサを用いて機械学習および戦略判断を行うコンピュータからその挙動を制御する.

本論文では, インタビュー対象者からより多くの情報を引き出すために, 発話ターンごとに態度推定を行い, 推定結果に基づき質問戦略決定を行う方法について検討を行う.

インタビューログデータから記事やハイライト動画を作成するための技術・検討に関しては，[石原 17] で述べる.

連絡先: 長澤 史記, 東京工業大学大学院情報理工学院 情報工学系 情報工学コース, 神奈川県横浜市緑区長津田町 4259 J2-53, TEL:045-924-5214

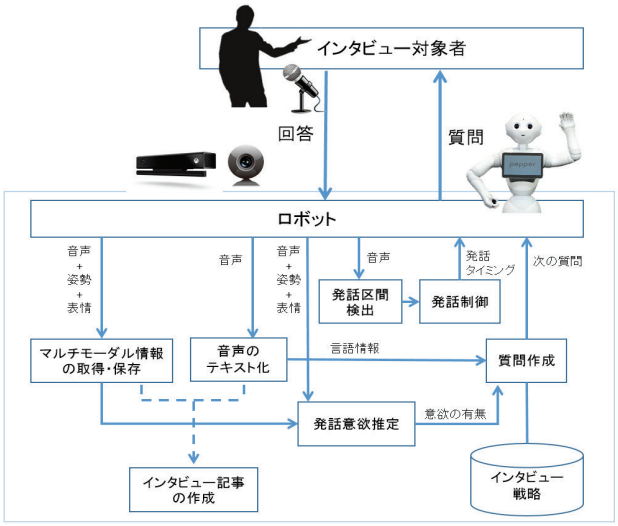

図 1: システム構成

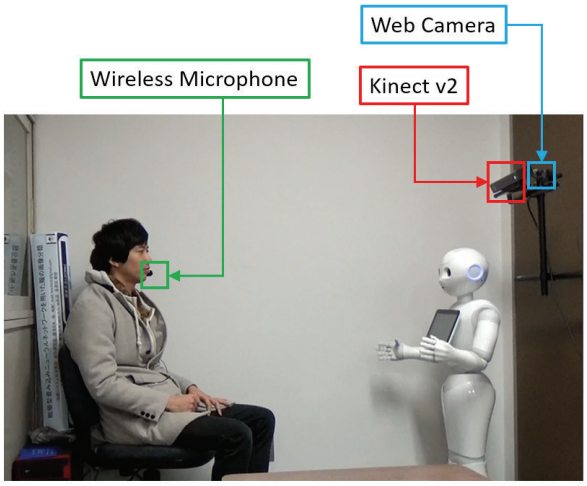

図 2: インタビュー実験のようす

3.1 マルチモーダル情報の取得

インタビュー対話を行っている間, ロボットの後方に設置した Logicool 社製 Web カメラを用いて対象者の顔画像を 30fps で撮影し, Kinect v2 から対象者の姿勢の情報を取得する. Web カメラと Kinect は近くに設置し, インタビュー対象者を同じ方向から撮影する. 取得した顔画像に対して seeingmachine 社製顔認識ライブラリである faceAPI を用いて顔認識を行い, 顔の向きや視線方向の情報を取得する予定である.

音声については, Web カメラ内臓のマイクから対話全体の音声を取得するほか, Shure 社製の指向性マイクを対象者が装着し, 回答発話を行う際の音声を取得する.

3.2 発話区間検出と発話制御

システムは音声から Julius による発話区間検出を行い, 発話終了を検出すると, インタビュー戦略に則った次の質問を行う. この際, 発話終了とみなすタイミングは, 音が途切れるまでの連続区間だけではなく, 息継ぎやいいよどみなどを含めた一つながりの発話全体の終了を考える [中野 15]. これにより検出された終了タイミングから, 次の質問に移ってよいかを判断する発話制御を行う予定である.

3.3 発話意欲の推定と戦略判断

測定した姿勢や音声などのマルチモーダル情報を基に機械学習を用いて, 対象者の回答発話が意欲的であるか, 多く語っていたかなどといった発話態度の推定を行うことを想定する.

推定された発話態度ラベルに応じて以下の行動を行う.

- 意欲的かつさらに情報を引き出すことが可能なら, 直前に質問した話題についてより深く掘り下げた質問を行う
- 意欲的であるが回答が少ない場合には, より具体的な質問をして回答を促す.
- 意欲が小さい場合には, 質問する話題の変更を行う.

4. 実験

4.1 インタビュー対話実験

態度推定による適応的なインタビュー対話のための初期検討として, インタビュー対話を行いながらインタビュー対象者のマルチモーダル情報を収集するロボットシステムを開発し, 5 人にそれぞれ 1 セッションずつ, 計 5 セッションのインタビュー実験を行った.

対象者は図 2 のように椅子に座ってロボットと対面した状態で, 自身の研究活動に関する 15 項目についての質疑応答を行った. 質問内容とその順番は事前に用意した内容で, すべての実験で同じ質問を使用した. 今回の実験に使用した質問内容のリストを表 1 に示す.

表 1: 実験に使用した質問リスト

順	内容
1.	現在の研究テーマ
2.	研究を始めた時期
3.	研究テーマ選択の動機
4.	研究をしていて 楽しかったこと
5.	研究をしていて つらかったこと
6.	研究の魅力
7.	その研究にはどのような応用があるか
8.	研究のほかに興味のあること
9.	以前の研究
10.	以前の研究ではどんな成果を得たか
11.	今の研究とどちらが楽しいか
12.	それはなぜ？
13.	どんな趣味を持っているか
14.	私生活と研究との両立
15.	対話の感想

これによって得られたマルチモーダル情報を用いて態度推定を行うために有用な特徴量を調べるために, T 検定による分析を行った.

4.2 検討した特徴量

本実験では, マイクを用いて録音した音声について得られた音声特徴量と, Kinect によって得られた姿勢特徴量について検討した.

4.2.1 発話区間

インタビュー実験の終了後, 質問に対応する 1 応答毎の発話区間について人手によるアノテーションを行った. 1 質問に対

表 2: 積極-非積極間での T 検定結果

	特徴量 (姿勢)	特徴量 (音声)
5%以下	肩 Y 座標平均 肩 Z 座標平均 肘 Y 座標平均（片側） 肩ノルム平均（片側）	発話長 最大ピッチ 最小エネルギー MFCC 平均
2%以下		最小ピッチ 平均ピッチ

する回答発話を 1 発話区間とし, 発話区間の時間を発話長として特徴量に用いた.

4.2.2 姿勢

本実験では, インタビュー対象者は椅子に座った状態であるため, Kinect を用いて得られた骨格情報のうち上半身の情報である頭, 肩, 肘, 手の位置について検討した. 各部位について測定した (x,y,z) 座標のそれぞれの測定値を使用した. さらに, 原点からのノルム $\sqrt{x^2+y^2+z^2}$ を使用したほか, 肩, 肘, 手の左右がある部位については左右の値を加算したものも用いた. これらの特徴量について, 発話区間毎に平均と分散を計算した.

4.2.3 音声

ユーザーに装着したマイクから録音した音声について, 各応答の発話時間長の他, Speech feature extraction code [*1] を用いてピッチ, エネルギー, MFCC を求めた. これらの特徴量のそれぞれについて, 1 発話区間当たりの最大, 最小, 平均を計算した.

4.3 T 検定による特徴量の分析

各応答発話について,

1. インタビュー対象者が積極的に発言を行っているか (5 段階評価)
2. この応答を受けて, システムが次に行う行動:
   - 話題を掘り下げる質問を行う
   - 質問を補足して回答を促す
   - 別の話題に切り替える

の二種類のアノテーションを行った.

本論文では, 二つのアノテーション結果のうち, 発話への積極性についての分析を行った. 各発話の積極性 A_n について 1 から 5 の五段階でのアノテーション結果について, 意欲が高い ($A_n > 3$) 発話を A 群, 積極性が低い ($A_n < 3$) 発話を B 群とした T 検定を行い, 有効な特徴量の検討を行った. 個別のセッションごとに検定を行ったほか, 全セッションのデータを結合したものについても同様に検定を行い, 有意差の得られる特徴量を検討した. 次に行うべき発話に関してのアノテーション結果については今後の課題とした.

5. 結果

T 検定の結果, セッション間で共通して有意差が見られた特徴量を表 2 に示す. 表 2 より, 音声と姿勢に関わるいくつかの特徴量について有意差が見られた.

∗1 Speech feature extraction code, http://groupmedia.mit.edu/data.php

姿勢に関する特徴量については, 座標およびノルムの発話区間平均について, 共通して有意差が確認された. 一方で分散については, 対象者によっては有意差が認められる特徴量が見られたが, 有意差の現れる部位やその方向について個人差が大きく, 複数の対象者について共通した特徴量は認められなかった. また, 肘や肩ノルムなどの特徴量については, 左右合わせた数値ではなく左右どちらかの数値のみで有意差が見られた. これが左右どちらであるかはインタビュー対象者によって違いが見られた. 音声特徴量については, 有意差のみられた特徴量はどれも意欲が高い場合に高い数値を示した. 姿勢特徴量については, 意欲が高い発話中の座標の平均は, 意欲が小さい場合と比べて小さい値を示した.

6. 結論

本論文では, インタビュー対象者の態度を推定することにより, 対象者の対話への意欲を高めつつ, より多くの情報を引き出すインタビュー対話ロボットシステムの開発に向けた初期検討として, マルチモーダル情報を用いた態度推定に有用な特徴量の検討を行った.

この検討を行うために, 対象者のマルチモーダル情報をインタビュー対話中に取得するロボットシステムを構築した. インタビュー実験を行い, 得られたマルチモーダル情報に対して, 発話の積極性との関係について T 検定を用いた分析を行った結果, 音声特徴量については, 発話長とピッチの最大と最小と平均, および MFCC とエネルギーについて有意差が存在することを確認した. 姿勢特徴量についてはインタビュー対象者によって左右の違いが見られたが, 肩や肘の座標に共通して有意差が見られた.

今後は, 応答を得た際の次に行う行動についてのアノテーションデータについて検定を行う. 検定の結果を踏まえて特徴量を選択し, 機械学習を行い, 態度推定モデルを構築する. また, 態度推定に応じたインタビューを行うための対話戦略構築を行う予定である.

参考文献

[Saito2015] Saito, Naoko and Okada, Shogo and Nitta, Katsumi and Nakano, Yukiko and Hayashi, Yuki: Estimating User's Attitude in Multimodal Conversational System for Elderly People with Dementia, 2015 AAAI Spring Symposium Series(2015)

[岡田 16] 岡田将吾, 松儀良広, 中野有紀子, 林佑樹, 黄宏軒, 高瀬裕, 新田克己: マルチモーダル情報に基づくグループ会話におけるコミュニケーション能力の推定, 人工知能学会論文誌 (2016 年 08 月)

[石原 17] 石原卓弥, 長澤史記, 岡田将吾, 新田克己: インタビュー対話における重要シーン抽出のための言語・非言語特徴量の分析, 人工知能学会全国大会 (2017)

[千葉 14] 千葉祐弥, 伊藤彰則: ユーザの対話意欲を考慮したユーザプロファイリング対話システムのためのインタビュー対話の分析, 電子情報通信学会技術研究報告 (2014)

[中野 15] 中野幹生, 駒谷和範, 船越孝太郎, 中野有紀子：対話システム (自然言語処理シリーズ), コロナ社 (2015)