

グループディスカッション参加ロボットの頭部動作決定機構の検討

Consideration of Group Discussion Participating Robot Autonomously Performing Head Movement

木村 清也 *1

Seiya Kimura

黄 宏軒 *2*3*4

Hung-Hsuan Huang

桑原 和宏 *2

Kazuhiro Kuwabara

*1立命館大学大学院情報理工学研究科

College of Information Science and Engineering, Ritsumeikan University

*2立命館大学情報理工学部

Graduate School of Information Science and Engineering, Ritsumeikan University

*3理化学研究所革新知能統合研究センター

Center for Advanced Intelligence Project, RIKEN

*4京都大学大学院情報学研究科 知能情報学専攻

Department of Intelligence Science and Technology, Graduate School of Informatics, Kyoto University

In recent years, Japanese companies are seeking employees with high communication skills. Therefore, there are companies that adopt group discussion at the time of job selection recruitment. Therefore it is thought that by raising communication skills the possibility of succeeding in job hunting can be increased. Repetitive practice is thought to be effective for improving communication skills, while a corresponding partner is required to do iterative practice. However, preparing partners is not easy for many students. We are developing a training system for group discussion with virtual agents and robots. In this paper, we propose for robots participating in group discussion.

1. はじめに

近年、日本企業は社員に対してコミュニケーション能力を求める傾向にあり、その能力を評価するために就職採用選考時にグループディスカッションを取り入れる企業が存在する。したがってコミュニケーション能力を高めることによってより就職活動に成功する可能性を高めることができると考えられる。コミュニケーション能力の向上には反復練習が有効であると考えられている一方で、反復練習を行うためには相応のパートナーが必要となる。しかし、パートナーを用意することは多くの学生にとって簡単なことではない。そこで我々は仮想エージェントやロボットによるグループディスカッションの訓練システムの開発を進めている。ロボットがグループディスカッションに参加するためには、人間にとって自然な振る舞いをしてることが求められる。現在は振る舞いの一つである頭部動作に着目し、グループディスカッション中に他の参加者の非言語行動、言語行動からロボットの頭部動作を決定するシステムの実現を目指している。このシステムは参加者の言語、非言語行動を各種センサーによってデータを収集し、収集されたデータを基に入力モジュール群によってデータの解釈と特徴量を抽出する。その特徴量を基に頭部動作決定部によってロボットの頭部動作を決定するシステムである。そしてこの論文ではシステムの中の機構の一つである頭部動作決定部の検討を行う。本稿は以下のように構成されている。第2章では関連する研究について紹介し、第3章で頭部動作決定システムの全体の構成について述べる。第4章でシステムの中で頭部動作を決定する機構を構築について述べ、最後に5章でまとめを行う。

2. 関連研究

就職採用面接の訓練のための仮想エージェント開発はこれまでも行われてきた [Baur 3, Chollet 5, Jones 6]。しかしこれらは1対1の面接訓練のための仮想エージェントである。複数人会話であるグループディスカッションに参加できるロボットを実現するためには、1対1の対話にはない課題を乗り越える必要がある。[David 2]では、複数のユーザーとエージェントとのインタラクションを実現する上での主要な課題について検討されている。また、これまでになされてきた複数のユーザーとエージェントとのインタラクションに関する研究の多くはエージェントやロボットが聴き手としての対話に参加するなど人間との立場が対等でない [Iolanda 4, Marynel 7]。一方我々の研究では、人間とロボットが対等な立場でグループディスカッションを行うことを目標としているため、それに向けたシステムの構築に取り組む必要がある。

3. 頭部動作決定システム

グループディスカッションに参加するロボットののための頭部動作決定システムについて述べる。このシステムはグループディスカッション中に他の参加者の言語行動と非言語行動からロボットの頭部動作を決定するシステムである。システムは大きく分けてセンサー群、入力理解モジュール群、頭部動作決定部の3つの要素に分けられる。図2にシステムの構成を示す。

(a) センサー群

各参加者および実験環境に取り付けられた各種センサーからグループディスカッション中の参加者の非言語行動、言語行動データを収集する。

(b) 入力理解モジュール群

このモジュール群ではセンサー群によって得られた生デー

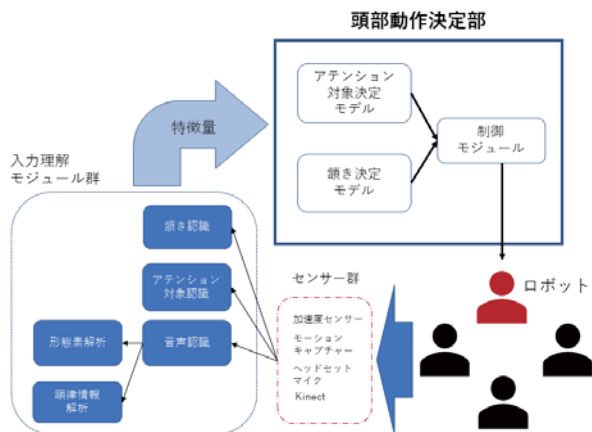


図 1: 頭部動作決定システム構成

タを抽象度の高いデータに変換し、それらを基にアテンション対象特徴量、発話ターン特徴量、動作特徴量、言語特徴量を抽出し、頭部動作決定部に送信する。

(c) 頭部動作決定部

入力理解モジュール群によって抽出された特徴量からロボットが行う頭部動作を決定し、ロボットに送信する。

この論文では (c) 頭部動作決定部の構築について述べる。

4. 頭部動作決定部の構築

頭部動作決定部の構築を行う。頭部動作決定部では入力理解モジュール群によって抽出された特徴量をもとにロボットの以下の2つに関する頭部動作を決定する。1つ目はロボットがアテンションを向ける対象を決定する。ここでは注視方向と頭部方向の組み合わせをアテンションとして扱い、ロボットが向けるべき対象を決定する。2つ目はロボットが頷くタイミングを決定する。これらの頭部動作を決定するためにこの機構はアテンション対象決定モデル、顔き決定モデル、そして制御モジュールによって構成されている。アテンション対象決定モデルと顔き決定モデルはロボットの頭部動作の決定を行い、結果を制御モジュールに出力する。制御モジュールは各モデルによって出力された結果を整理し、実際にロボットが行う頭部動作の指令をロボットに送信する。

4.1 対話実験データの収集

アテンション対象モデルと顔き決定モデルを生成するために対話収集実験を行った。

4.1.1 実験概要

実験手順は、Fumio らに従ったメガネ型イトラッカーは使用しなかった [Fumio 11]。ディスカッションの議題は、日本企業の就職採用面接において頻繁に使用される「サバイバルタスク」型の課題を設定した。この課題は与えられた選択肢の中から提示された条件に基づいて順位づけを行う課題である。今回の実験では配布した資料の中に記載された15名の有名人の中から収益や集客を考慮し、最適だと思われる人物の順位をつけていく「学園祭に招待する有名人ランキング」を議題として設定した。また議論が活発となるよう、議論に取り組む前にそれぞれの参加者で課題について考える時間を5分間設けた。実験データは互いに顔見知りでない学生4人を1グループと



図 2: 実験環境

した10グループの合計40人分のデータによって構成されている。また、より実践的なグループディスカッションを行ってもらうため実験参加者は就職活動経験者、就職活動中の学生を対象とし、議論中の男女の比率の違いによる発言の優劣をなくするため、グループ内の男女の割合は、すべて同性か、異性の数が等しくなるように設定した。

4.1.2 実験環境

1.2m × 1.2m の正方形のテーブルの周りに座って議論を行う実験参加者の様子を、様々なセンサをとビデオカメラを用いて記録した。加速度センサー (ATR-Promotions TSND121) とオーディオデジタイザ (Roland Sonar X1 LE) に接続されたヘッドセットマイク (Audio-Technica HYP-190H) を参加者の頭に装着し、各参加者の顔を4台のウェブカメラ (Logicool C920) を用いて撮影した。

4.2 アテンション対象決定モデル

対話実験データからアテンション対象決定モデルを生成する。このモデルでは入力理解モジュール群によって抽出された特徴量をもとにアテンション対象決定モデルを用いてロボットがアテンションを向けるべき対象を決定し、制御モジュールに結果を出力する。ロボットはアクチュエータが人間と同じように機能せず、人間よりも自由度はるかに低く、眼球もない可能性もあることからこの研究ではロボットの視線方向や頭部の向きなどの全体的な組み合わせをアテンションとして扱う。アテンションを向ける対象となるのはロボットの視点から見て、左側に座っている参加者、テーブルの反対側に座っている参加者、右側に座っている参加者とさらに配布した机上の資料 (今回の実験では有名人リスト) の4つのである。モデルを生成するためにアテンション対象に関する特徴量が15個、発話ターンに関する特徴量23個、顔きに関する特徴量36個、動作に関する特徴量が18個の計93種類の特徴量を用いてモデル生成した。また、モデルは「ロボットが発話をしている場合 (Speaking)」 「他の参加者が発話をしている場合 (Listening)」 「参加者全員が発話をしていない場合 (Idling)」 の3種類のシチュエーションに分け、非線形SVM (Support Vector Machine) を用いてコストパラメータ $C = 10.0$ RBF カーネルのパラメータ $\gamma = 0.01$ によって生成した。評価には Leave-one-person-out 方を用い、F値は0.4~0.6程度となった。表1は各シチュエーションごとのモデルの精度である。

表 1: アテンション対象モデルの分類結果

Situation	Precision	Recall	F-measure
Idling	0.537	0.536	0.536
Listening	0.566	0.566	0.566
Speaking	0.453	0.450	0.452

4.3 領き決定モデル

対話実験データから領き決定モデルを生成する。このモデルでは入力理解モジュール群によって抽出された特徴量をもとに領き決定モデルを用いてロボットが領きを行うべきかを判定する。判定は他者の発話終了時毎に行われ、その結果を制御モジュールに出力する。このモデルでは、発話の言語特徴量や、発話中の話者やそれ以外の参加者の非言語行動から特徴量を抽出しモデルを生成する。モデルを生成するためにアテンション対象に関する特徴量が 13 個、発話ターンに関する特徴量 6 個、韻律に関する特徴量 4 個、動作に関する特徴量が 7 個、言語特徴量が 4 個の計 34 個の特徴量を用いた。モデルの生成には非線形 SVM を用いてコストパラメータ $C = 20.0$ RBF カーネルのパラメータ $\gamma = 0.1$ によって生成した。評価には Leave-one-person-out 方を用い、その結果 F 値は 0.796 となった。

4.4 制御モジュール

このモジュールではアテンション対象決定モデルと領き決定モデルによって出力された結果をもとに実際にロボットが行う頭部動作の指令を送信する。特徴量をもとにアテンション対象モデルによって出力された結果は、制御モジュールによってアテンション対象の平均継続時間である 9.6 秒間の過去の出力結果が蓄積される、そしてその 9.6 秒の間で最も出力が多かった対象をアテンションを向ける対象として決定しロボットに送信する。また、グループディスカッションの開始から経過した時間が 9.6 秒に満たない場合は開始からその時点でまでで最も多かった出力結果をアテンションを向ける対象として決定しロボットへ送信する。領きモデルでは特徴量をもとに他者の発話の終了時に領きを行うべきかが判定され、領きを行うべきと判定された場合は制御モジュールが実験参加者の領きの平均継続時間である 1.3 秒間の領きを行う指令をロボットに送信する。アテンション対象の切り替えと領きが同時に発生した場合は領きが優先される。また、ロボットが領いている間はアテンション対象の切り替えは行わないが、領き行っている間もアテンション対象モデルによる結果の出力と制御モジュールへの蓄積は行い、領きが終了した時点で領きを行っていた時間も含めた 9.6 秒の間で最も出力が多かった対象をアテンションを向ける対象として決定しロボットに送信する。領き行っている間に新たに領きを行うべきと判定された場合、その時点から 1.3 秒間の領きを行う指令をロボットに送信する。

5. 終わりに

この論文ではグループディスカッションに参加するロボットの頭部動作システムを構築するために必要な機構の一つである頭部動作決定部の提案を行った。頭部動作決定部は入力理解モジュールによって抽出された特徴量を基に、頭部動作決定部が頭部動作を決定しロボットに出力する。今後は入力理解モジュール群の実装も進めることで、リアルタイムに頭部動作を決定するシステムの構築を目指す。

参考文献

- [Baur 3] Tobias Baur, Ionut Damian, Patrick Gebhard, Kaska Porayska-Pomsta, and Elisabeth Andre : A Job Interview Simulation: Social Cue-based Interaction with a Virtual Character, In 2013 International Conference on Social Computing, pp. 220-227 (2013)
- [Jones 6] Hazael Jones, Mathieu Chollet, Magalie Ochs, Nicolas Sabouret, and Catherine Pelachaud : Expressing social attitudes in virtual agents for social coaching, Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems, pp. 1409-1410 (2014)
- [Chollet 5] Mathieu Chollet, Magalie Ochs, and Catherine Pelachaud : Mining a Multimodal Corpus for Non-Verbal Signals Sequences Conveying Attitudes. In The 9th edition of the Language Resources and Evaluation Conference, pp. 3417-3424 (2014)
- [Traum 2] David Traum : Issues in Multiparty Dialogues. In Advances in Agent Communication, International Workshop on Agent Communication Languages In Proceedings of the 16th International Conference on Multimodal Interaction, pp. 201-211 (2003)
- [Fumio 11] Fumio Nihei, Yukiko I. Nakano, Yuki Hayashi, Hung-Hsuan Huang, and Shogo Okada : Predicting Influential Statements in Group Discussions using Speech and Head Motion Information, pp. 136-143 (2014)
- [Leite 4] Iolanda Leite, Marissa McCoy, Monika Lohani, Daniel Ullman, Nicole Salomons, Charlene Stokes, Susan Rivers, and Brian Scassellati : Emotional Storytelling in the Classroom: Individual versus Group Interaction between Children and Robots. In 10th ACM/IEEE International Conference on Human-Robot Interaction, pp. 75-82 (2015)
- [Vazquez 7] Marynel Vazquez, Elizabeth J. Carter, Braden McDorman, Jodi Forlizzi Aaron Steinfeld, and Scott E. Hudson : Towards Robot Autonomy in Group Conversations: Understanding the Effects of Body Orientation and Gaze, Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction, pp. 42-52 (2017)