2G1-3

グループディスカッションに参加するロボットにおけるアテンション対象モデルの提案

Proposal of the Attention Model for a Robot in Group Discussion Context

木村 清也 *1 Seiya Kimura　張 琪 *1 Qi Zhang　黄 宏軒 *2 Hung-Hsuan Huang　岡田 将吾 *3 Shogo Okada　林 佑樹 *4 Yuki Hayashi　高瀬 裕 *5 Yutaka Takase

中野 有紀子 *3 Yukiko Nakano　大田 直樹 *2 Naoki Ohta　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東京工業大学大学院総合理工学研究科
Graduate School of Science and Engineering, Tokyo Institute of Technology

*4大阪府立大学現代システム科学域知識情報システム学類
College of Sustainable System Science, Osaka Prefecture University

*5成蹊大学理工学部情報科学科
Department of Computer and information Science, Seikei University

In recent years, companies are seeking for communication skill from their employers. Companies that use group discussions in employer recruitment to evaluate communication skills are therefore increasing. However, the opportunity for improving communication skill in the group discussion context is limited. In order to solve this issue, our ongoing project is aiming to build a communication robot that can participate group discussion, so that its users can practice group discussion over and over. In this paper, we propose the models directing the robot ' s attention toward the other participants in three situations: speaking, listening, and idling. First, we gathered a data corpus of the discussion of four-people groups. Then the multimodal features including voice prosody, head movements, and speech turn were extracted from the corpus and were used to train the models with random forest algorithm.

1. はじめに

近年の日本企業が社員に求める能力としてコミュニケーション能力があげられる．コミュニケーション能力は人が相手と意思や思考の伝達を円滑に進めるために必要な能力であり，そのため近年では約 40 %の企業が就職採用選考おいて，就職活動者のコミュニケーション能力を評価するためにグループディスカッションを取り入れている [Hirano 2011] ．

企業から高いコミュニケーション能力の評価を得て，ひいては内定獲得の可能性を高めるのに，グループディスカッションの反復練習は有効であると考えられる．本研究は，グループディスカッション練習システムの一環として，グループディスカッションに参加できるコミュニケーションロボットの開発を目指す．グループディスカッションにロボットが参加することにより，参加者が不足している場合でも練習を行うことが可能となる．またロボットが他の参加者の所作を評価・分析し議論中にロボット自身が会話操作をし参加者に話を振ることで，コミュニケーション能力の改善に役立てることが期待される．

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，ロボットが複数人と会話する状況において人間にとって自然な振る舞いをしていることが機能として求められる．これを達成するために，適切なロボットの行動モデルを作成する必要がある．そこで，実際人間がグループディスカッションに参加している様子を収録した実験データからから参加者が行うコミュニケーション行動を分析し，グループディスカッション参加型ロボット作成の基礎的検討を行う．本論文では，ロボットのアテンションを向ける対象を決定するモデルとして,「ロボットが発話をしている」,「他の参加者が発話をしている」,「全員が発話をしていないとき」の 3 種類のシチュエーションごとに分けたものを提案する．各モデルは，それぞれの状況に応じて，実験データから抽出したマルチモーダル特徴量に対して RandomForest を用いて学習した．

連絡先: 黄 宏軒，立命館大学情報理工学部，滋賀県草津市野路東 1 － 1 － 1 ，hhhuang@acm.org

2. 関連研究

複数人対話における非言語の分析については熊野ら [熊野 2015] が内向き・外向きの双方向を撮影するウェアラブルカメラを装着し個人の特性や内部状態を表す重要指標の一つである視線行動をほぼ自動で分析する方法を提案した．石井ら [石井 2015] は，視線と呼吸動作のマルチモーダル情報を用いて次話者の予測をするモデルを構築し，視線と呼吸情報の両方を用いた場合の予測性能が最も高いことを示した．これらの研究では対話における非言語行動のパターンや会話イベントの推定あるいは予測が行われている．しかし，それらの情報を特徴量として分析しエージェントを実際に複数人対話へ参加させることを目的とした研究が行われている．本研究もこちらに属する

吉野ら [吉野 2015] は会話参加者の顔向きや発話量から会話に

おける優位性を逐次的に推定し，推定結果や会話内容をもとにエージェントが対話に介入する対象とタイミング，内容を決定するシステムの提案を行った．小山ら [小山 2015] は，複数人対話において参加者の振り向き，および発話タイミングを分析し，エージェントの複数人会話における聞き手としての振る舞いを検討した．両者とも対話におけるシチュエーション，が明確でないが，本研究ではグループディスカッションに参加するロボットの実現のため，シチュエーションを明確し，対話参加者の非言語動作の分析を行う．

3. 対話実験データの収集

3.1 対話実験の実施

人間同士での会話においてアテンション対象の分析に必要なデータを収集するために，対話収集実験を行い，グループディスカッション対話コーパスを構築した．被験者 4 名から構成されるグループに課題を与え，グループディスカッションを行ってもらい，各被験者の非言語行動を各種機材によってマルチモーダル情報として記録した．

就職活動にて行われるグループディスカッションのテーマは「自由討論」,「インバスケット」,「ケーススタディ」,「ディベート」があることが指摘されている [林 2015]．またこれらのテーマの中でも特にが就職活動のテーマとして設定されることが多い,「(a) ケーススタディ」の議題を 2 題と「(b) インバスケット」の議題を 1 題，の計 3 題を今回の収集実験におけるグループディスカッションのテーマとした．インバスケット型の議題は絶対的な正解がなく，もっとも重要なものを選ぶ過程を観察できる．今回の収集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（b）では設定されている目標をクリアするために参加者全員で最良の方策を考え出す過程を観察できる．今回の収集実験では外国人の友人が夏に一泊二日で来日するという仮定のもとで，友人が喜ぶと思われる旅行の計画を立てる「外国人の友人のおもてなし計画」と資料をもとに学園祭の出店内容と出店場所を決める「学園祭の出店計画」を議題として設定した．

実験参加者は互いに顔見知りでない学生 4 人を 1 グループとした 15 グループの合計 60 人（男性 48 人女性 12 人）である．また，就職活動経験者，就職活動中の学生に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，大学院生以下，大学 3 年生以上の学生を被験者の対象とした．実験を実施にあたり男女の比率の違いによる発言の優劣をなくすため，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（男 4 人グループが 9 組，男女 2 : 2 のグループが 6 組）．セッションを行う順番にも違いが出てくることを想定して，実験ごとにセッションの順番を変更する操作も施した．また，各被験者に白紙の A4 用紙を配布し，自由に使ってよいと伝えた．各課題の制限時間は 15 分とし，より確実に課題に取り組んでもらうために「学園祭に招待する有名人ランキング」,「学園祭の出店計画」は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 3 分設けた．制限時間を示すタイマーは，各参加者が見ることが出来る位置に設置し，制限時間の終了はブザーを鳴らすことによって知らせる．

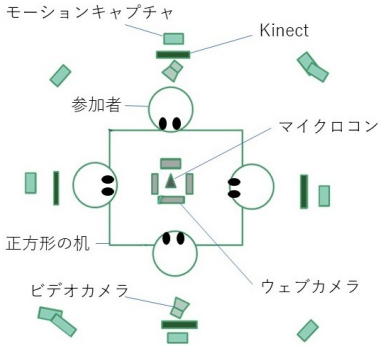

図 1: データ収集環境のレイアウト

3.2 実験環境

被験者は上記で示した議題について，図 3.1 のような環境で対話を行った．以下で対話中のデータを記録・収録するため使用された機材を説明する．

音声記録は（SONAR X1 LE　Roland 製）のヘッドセットマイクを使用し，ソフトウェアにて録音．さらに，SONAR X1 LE だけでなく，発話者の方向を特定し適応的にノイズを低減することができる Microcone（devaudio 製）も使用し音声収録をおこなった．

映像から参加者を観察し，注視方向を知るために参加者一人ずつの表情を撮影するため Web カメラ (C920 Logicool 製)4 台を机に設置した．さらに，対話を撮影するため 3 台のビデオカメラを設置した．2 台のビデオカメラで 2 人ずつの参加者の撮影を行った．そして，3 台目のビデオカメラはタイムスタンプが映っているモニターを撮影することで全てのデータの同期を実現．

図 2: 被験者に装着する機材

4. マルチモーダル特徴量の抽出

対話収集実験によって得られたデータから，特徴量を抽出し教師データを作成するために，実験で実施した 3 種類の課題のうち，重要なものを選ぶ過程を観察できる，インバスケット型課題である“学園祭に招待する有名人ランキング”のコーパスを分析対象とした．映像データと音声データが一部欠損していたため最終的に 10 グループ分のデータを用いた．グループディスカッション参加者 4 人のうち 1 人に着目し，その参加者の視点からそれ以外の他の 3 人のコミュニケーション行動を特徴量として 0.1 秒刻みで抽出した．1 章にて述べた 3 種類のシチュエーションである「ロボットが発話をしている時」

「ロボットが発話をしておらず，他の参加者が発話をしている時」,「ロボットを含む参加者全員が発話をしていないとき」をそれぞれ「speaking」「listening」「idling」と呼称する．マルチモーダル特徴量として，発話ターン特徴量，アテンション特徴量を抽出した．

4.1 発話ターン特徴量

ヘッドセットマイクによって収録された参加者それぞれの音声記録を用いて，音声分析ツールである Praat[*2] を利用し参加者の音声から有音と無音の識別を行い発話区間データの収集を行った．そのデータを利用し発話ターン特徴量を抽出する以下は発話ターン特徴量についてまとめたものである．

発話状態： 参加者の発話の有無.

発話回数： グループディスカッション開始からその時点までの発話の累計.

発話長： 発話の長さ

発話率： グループディスカッション開始からその時点までに発話を行っていた時間の割合を計算する.

各状態になってからの経過時間： 「idling」「listening」「speaking」の各状態になってから経過した時間.

4.2 アテンション特徴量

web カメラによって得られた実験映像から，アノテーションツール ELAN[*1] を利用し，各参加者のアテンション対象を抽出する．アテンション対象のアノテーション行った際のラベルの名称と定義は「table ラベル」アテンションを下（テーブル）に向けている「front ラベル」アテンションを対面に座っている人物に向けている.「right ラベル」アテンションを右側に座っている人物に向けている.「left ラベル」アテンションを右側に座っている人物に向けている.「away ラベル」アテンションを上記の 4 つ以外に向けている．この定義に基づいて作成されたアノテーションデータを集計したもの表に示す．table ラベルが多くなっているのはグループディスカッション参加者が配布した白紙の A4 用紙に書き込んでいる，あるいは配布資料を読んでいることが多いことが原因として考えられる．また，away ラベルはほとんどの参加者でアノテーションされることがなく，サンプルが圧倒的に少なくなってしまったため，今回の研究では最も近似したアテンション対象にラベルにアノテーションをし直した．

表 1: 全参加者のアノテーションデータ

ラベル	個数	平均継続時間	最大継続時間	最小継続時間
table	1715	17.7 秒	359 秒	0.38 秒
front	1075	2.6 秒	31 秒	0.18 秒
right	662	2.3 秒	30 秒	0.18 秒
left	642	2.1 秒	25 秒	0.11 秒
away	4	1.8 秒	4 秒	0.10 秒

∗2 http://www.fon.hum.uva.nl/praat/
∗1 http://tla.mpi.nl/tools/tla-tools/elan/

作成されたアノテーションデータをアテンション特徴量として抽出する際に，着目している参加者から見た方向の人物に変換する．以下はアテンション特徴量についてまとめたものである．

アテンション対象： 各参加者のアテンション対象.

アテンションが着目している参加者へ向けられた割合： 各参加者のグループディスカッション開始からその時点までに着目している参加者にアテンションを向けていた割合を計算する.

4.3 韻律特徴量

4.1 節にて得られた発話区間の情報をもとに各発話の韻律情報の特徴を抽出する．特徴抽出には Praat を用いて基本周波数をの解析を行う．基本周波数はピッチ特性と大きな相関をもち実質的に声の高さとして扱われている．そのため厳密には一致しないが本研究では基本周波数をピッチ特徴量として扱う．また，今回声の大きさを特徴をしてとらえるためにインテンシティレベルに関する特徴量の抽出を行う

ピッチ： 各参加者の発話区間におけるピッチレベル．単位は「Hz」.

平均ピッチとの差： 各参加者の発話区間における平均ピッチを計算し，その数値とその地点におけるピッチレベルとの差を計算する.

インテンシティ： 各参加者の発話区間におけるインテンシティレベル，単位は「d B」

平均インテンシティとの差： 発話区間における平均インテンシティを計算し，その数値とその辞時点におけるインテンシティレベルとの差を計算する.

5. アテンション対象モデルの生成

前章にて抽出した特徴量をもとに 3 種類のシチュエーションごとにアテンション対象モデルの生成を行う．着目している参加者以外の 3 人の特徴量セットを説明変数，着目している参加者のアテンション対象を目的変数として将来リアルタイムでアテンション対象を決定することを想定し，分類速度が高速かつ精度が高い RandomForest アルゴリズムを用いて学習を行う．生成する木の数を表すパラメータである I を 100 とする．

各モダリティのモデル精度への貢献度を検証するために，3 種類の特徴量の 7 種類の組み合わせでモデル精度の比較を行う．アテンション特徴量，発話ターン特徴量，韻律特徴量をそれぞれ A，S，P，と略記する．なお，idling 状態モデルについては，グループディスカッション中の参加者全員が発話を行っていない状態であるため，発話ターン特徴量である，発話状態は全員が無の状態であるため，特徴量には含まないとする．同じく発話長についても特徴量には含まないとする．また，発話区間内で抽出される韻律特徴量は全員が 0 である．従って idling モデルのみ，韻律特徴量を含まない (1)(2)(4) の特徴量セットを用いてのモデルの生成，および特徴量セットごとの精度比較を行う．各モデルの評価には leave-one-group-out 法を用いる．

（1）A：アテンション特徴量
（2）S：発話ターン特徴量
（3）P：韻律特徴量
（4）A+S：アテンションと発話行動

（5）A+P：アテンションと韻律
（6）S+P：発話ターンと韻律
（7）A+S+P：アテンションと発話行動と韻律

6. 結果

得られた結果を各モデルごとに図 3 に示す．また，各モデルにおいて最も精度が高かった特徴量セットの F 尺度を表 2 に示す． speaking 状態モデルでは韻律情報のみを特徴量として

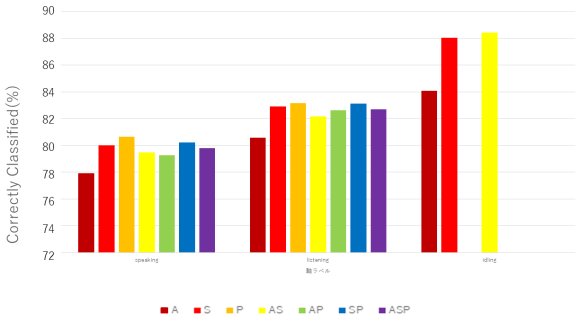

図 3: 分類精度比較

表 2: 最も精度が高かった特徴量セットの F 尺度

モデル	特徴量セット	分類精度（%）	F 尺度
speaking	P	80.7	0.73
listening	P	83.2	0.72
idling	AS	88.4	0.84

モデルを生成した場合，分類精度が 80.7 %と最も高く，F 尺度は 0.73 となった．listening 状態モデルは韻律情報のみを特徴量として生成されたモデルの分類精度が最も高く 83.2 %となり，F 尺度は 072 となった．idling 状態モデルはアテンションと発話行動のセットを組み合わせたを特徴量としてモデルを生成した場合に分類精度が 88.7 %と最も高く，F 尺度が 0.84 となった．

韻律情報が存在する speaking 状態と listening 状態においては共に韻律情報のみを特徴量としたモデルの精度が最大となった．これによってグループディスカッション参加者は他者の今回の韻律情報として抽出した声の高さ，あるいは大きさによってアテンション対象を決定している可能性が高いことが示唆された．韻律情報が存在しない idling 状態では，他の状態とは違い複数の特徴量セットを組み合わせて生成したモデルの精度が最大となり，グループディスカッション参加者は全員が沈黙している場合，モダリティーの異なる複数の要素によってアテンション対象を決定している可能性が高いことが示唆された．

7. 終わりに

本論文では，グループディスカッションに参加できるロボットの実現に向かってロボットのアテンション対象決定モデルの提案をした対話収集実験からグループディスカッション参加者のコミュニケーション行動をマルチモーダルな視点で分析することによって，ロボットがグループディスカッションに参加した際に他の参加者の注視対象と発話の有無からアテンション，発話行動，韻律のマルチモーダル特徴量に対して RandomForest 学習を行い，F 尺度 0.8 前後の結果を得た．
今後は，今回の研究ではできなかったケーススタディ型のグループディスカッションについての分析を行う必要がある．また，今回の研究ではモデルの生成を行ったがロボットをグループディスカッションに参加させるため，今回手動でアノテーションを行っていた部分を自動化するシステムおよびグループディスカッション中にリアルタイムでアテンション対象を決定するシステムの構築や実際のロボットへの実装についても進めていく予定である．

参考文献

[林 2015] 林佑樹, 二瓶芙巳雄, 中野有紀子, 黄宏軒, 岡田将吾グループディスカッションコーパスの構築および性格特性との関連性の分析, 情報処理学会論文誌,Vol.56,No.4,pp1217-1227(2015)

[小山 2015] 小山大幾, 水本武志, 中村圭佑, 中臺一博, 今井倫太複数人会話における振り向き動作と発話動作解析,HAI シンポジウム (2015).

[Hirano 2011] K.Hirano: The problems of academic ability in the company: validate the elements of academic ability in hiring of new graduates,The Japanese journal of labour studies (2011).

[福原 2012] 福原佑貴, 中野有紀子複数ユーザ対会話ロボットとの多人数インタラクションにおける優位性の自動推定方法, 第 74 回情報処理学会全国大会講演論文集,Vol.74,No.1,pp.381–382(2012).

[吉野 2015] 吉野尭, 八城美, 高瀬裕, 中野有紀子会話エージェントによる優位性推定に基づくグループ会話への介入, 第 24 回人工知能学会全国大会論文集, Vol.29, pp.1-3(2015).

[石井 2015] 石井亮, 熊野史朗, 大塚和弘複数人対話における呼吸動作に基づく次話者予測,HCG シンポジウム (2015).

[熊野 2015] 熊野史朗, 大塚和弘, 石井亮, 大和淳司 : 集団的一人称視点映像解析に基づく複数人対話の自動視線分析,HCG シンポジウム (2015).