

グループディスカッションにおける議論状態の変化の検出

Detecting Discussion State Shifts in Group Discussions

二瓶 芙巳雄^{*1}
Fumio NIHEI

林 佑樹^{*1}
Yuki HAYASHI

中野 有紀子^{*1}
Yukiko NAKANO

^{*1} 成蹊大学理工学部

Faculty of Science and Technology, Seikei University

In recent years, group discussion is used in job interview process to assess the candidates' social abilities such as communication skill, leadership. However, the measurement of such abilities solely relies on interviewer's subjective judgments. Aiming at approaching this problem from multimodal corpus analysis, this study, first, collected conversation corpus for group discussion by four people. The collected data include speech, head motions, face images, eye gaze, bone data, and personality traits. Focusing on speech audio and head motions, we analyzed discussion state changes, and defined three kinds of discussion states: *ordered*, *unordered*, and *discontinuity*. By applying SVM, we created a discussion state recognition model, and found that the model performed very well for *ordered* and *discontinuity* states (F-measure were 0.85 and 0.76 respectively), but not very well for *unordered* state (F-measure 0.47).

1. はじめに

採用選考では、被面接者が一人の場合や複数の場合など、多様な面接が行われる。また最近では、テーマを与えて数人の被面接者にグループディスカッションをさせ、その様子から評価を決定することも頻繁に行われている。しかし、評価は面接者によって異なる場合もあるため、グループディスカッションにおける被面接者の行動に対して、客観的な評価を与えるとともに、被面接者に対して、よりよく評価されるための支援を行うには、データに基づくグループディスカッションの分析・モデル化が必要であると考えられる。

また、グループディスカッションをはじめとした議論では、互いに意見や提案を述べあう、意思決定における合意の確認をする、議論が停滞している、といった様々な状態があり、議論の支援やアドバイスをを行うシステムを実現するためには、システムが議論状態の変化を検知し、議論の状態に応じた支援を行う必要がある。

以上の課題に取り組むために、本研究では、(1) 採用選考におけるグループディスカッションを対象とし、多様な非言語行動データを含むグループディスカッション対話コーパスを構築する。(2) 次に、収集した対話データを分析し、グループディスカッションにおける議論状態の変化を検知するモデルを構築する。本研究では音声の非言語情報である韻律情報として、発話の有無と音圧、加えて加速度センサによる頭部位置の変化量に着目し、機械学習を用いてモデルを作成する。さらに、このモデルを収集した対話データに適用することにより、議論が活発に行われているか、停滞しているかを自動的に判別することができるかを検証する。

2. 関連研究

グループディスカッションやその参加者の分析・モデル化を目的として収集された対話コーパスとして、AMI [Jovanovic 2005]やMSC [Pianesi 2007]がある。AMIのコーパスでは4人

を1グループとし、各被験者に対して役割を与えた上で議論を行わせる。議論のテーマはテレビ用リモコンのデザインで、各被験者の役割はプロジェクトマネージャ、マーケティングの専門家、ユーザインタフェースデザイナー、工業デザイナーの4種である。その実験中の映像、音声を記録し、それらに対しダイアログアクト、受話者、発話の隣接ペア、視線方向のアノテーションを行っている。MSCコーパスではAMIと同じく4人を1グループとし、災害時など非常事態下を想定して複数の道具に対する優先順位を議論して決めさせるサバイバルタスクに取り組み、そこでのインタラクションを収集している。

また、グループディスカッションにおける言語・非言語行動を分析した研究として、大本ら [大本 2011] は、ファシリテーション行動に着目し、ファシリテータはどのような情報に基づきファシリテーション行動を選択しているのかを明らかにしている。この研究では、議論は“発散”と“収束”の状態を持ち、これを繰り返している [堀 2004] とし、ファシリテータは議論を発散あるいは収束させることによって議論を活性化していると述べている。

Jurafskyら [Jurafsky 1997] は、二者間の対話を対象とし、談話構造を音声の言語、非言語情報から自動で検出する方法を提案している。この研究では、談話に対して **Statement** や **Opinion** などのタグを手動で付与した後、韻律情報や音声認識によって得られた単語、単語の並び、また付与されたタグの並びを用いることによって、談話の種類を判定している。

本研究は、非言語情報から議論の状態を推定しようとするものであり、これらの研究とは目的が異なるが、コーパスの構築方法や議論の分析方法、推定に利用する非言語行動の選択などにおいて、上記の先行研究を参考にした。

3. 対話データの収集

3.1 対話収集実験の実施

グループディスカッション対話コーパスを構築するために、対話収集実験を行った。4名の被験者から構成されるグループに、採用面接におけるグループディスカッションで用いられるような課題に取り組み、各種機材によって各被験者の言語・非言語行動を記録した。

連絡先: 二瓶芙巳雄, 成蹊大学理工学部, 東京都武蔵野市吉祥寺北町3-3-1, us102080@cc.seikei.ac.jp

(1) 被験者

被験者は男性 32 名、女性 12 名、計 44 名の大学生である。被験者の学部や学科、学年は様々である。4 人の被験者を 1 つのグループにし、合計 11 のグループを形成した。各グループには、可能な限り就職活動経験者を 1 名入れ、男女同数となるようにし、またそのグループに知り合いが含まれないようにした。

(2) 実験課題

グループディスカッションの課題を選定するために、まず、採用面接におけるグループディスカッションでよく扱われる議題を事前に調査した。その結果、グループディスカッションの議題は、自由討論型、インバケット型、ケーススタディ型、そしてディベート型といったカテゴリに分類されることがわかった。今回の実験では、“インバケット型”の議題を 1 つ、“ケーススタディ型”の議題を 2 つ、計 3 つの議題を設定した。

インバケット型課題は、サバイバルタスクに代表されるような、複数の項目に対する優先順位をグループで議論して決める課題である。本研究では、サバイバルタスクをアレンジし、15 人の有名人のリストを与え、学園祭に呼びたい順位をランク付けする(学園祭有名人ランキング)という課題を設定した。

ケーススタディ型課題は、ある問題に対して、解決策の立案や問題点の分析を行うものである。本研究では、学園祭での出店計画(学園祭出店計画)、外国人を日本に招いたときの旅行プラン(外国人の友人おもてなし計画)の 2 つを設定した。

(3) 実験環境

実験を行うにあたり、専用の実験環境を用意した。実験環境を図 1 に示す。また、この環境において実験を実施している様子を図 2 に示す。実験スペースは一边を 450cm とした空間であり、その辺の端点と中点には支柱を立て、その間に黒のカーテンあるいは白い壁にて仕切った。各支柱にはモーションキャプチャシステム OptiTrack 用の赤外線カメラ、そして全体の様子を録画するビデオカメラを設置した。また、実験スペースの中心に一边 120cm のテーブルを配置し、そのテーブルを囲むように 4 人の被験者を着席させた。テーブルの中心には各被験者に向けてウェブカメラを 4 台設置し、被験者の顔映像を取得した。Kinect センサは被験者に近いほうが深度情報や骨格情報を取得しやすいことがわかった。そのため Kinect センサを被験者に近づけるために各被験者の背後に支柱を立てて Kinect センサを設置し、対面する被験者の計測を行った。

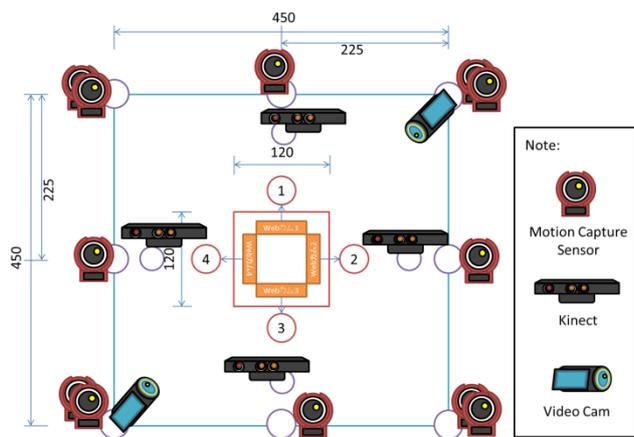


図 1 実験環境



図 2 実験の様子

(4) 手続き

対話収集実験に先立ち、各被験者に性格テストを実施した。性格テストの詳細は 3.2(7)節で詳述する。次に、各被験者に各種センサを装着してもらい(図 3)、各センサに必要なキャリブレーションを行った。各種センサについての詳細は 3.2 節で述べる。



図 3 各センサ装着の様子

各被験者グループに上述した学園祭有名人ランキング、学園祭出店計画、そして外国人の友人おもてなし計画の課題について議論してもらうことにより、3 会話を収録した。課題の順序による影響を除去するために、各グループにおける課題順はランダムにした。各課題の制限時間はインバケット型が 15 分、ケーススタディ型は 20 分とした。また各課題に取り組む前に、各自が他者と相談なしで考える時間をそれぞれ 2 分設けた。制限時間を示すタイマーは、各被験者が見ることが出来る位置 2 個所に配置してあり、また制限時間の終了時は、ブザーを鳴らして知らせた。

3.2 実験データの取得

44 名の被験者により 11 の被験者グループが構成され、各グループに対して 3 セッション実施したため、のべ 132 人分の言語・非言語データを収集することができた。

(1) 音声

被験者にはヘッドセットマイク(audio-technica : HYP-190H)を装着してもらい、各被験者の発話音声のデータを取得した。各ヘッドセットマイクからの入力 は、PC に接続されたオーディオインタフェース(Roland : UA-1000)に集積し、音声編集ソフト DigiOnSound5 を用いて録音した。これは複数チャンネルの同時録音に対応したソフトウェアである。実験終了後、各被験者の音声データを wav 形式に変換して保存した。一部ファイルの欠損のため、最終的に得られた音声ファイルは 128 ファイルとなった。

(2) 頭部動き量

被験者の後頭部に加速度センサ (ATR-Promotions : WAA-010) を取り付け、 x, y, z の 3 軸における加速度、角速度を 33msec ごとに計測することにより、各被験者の頭部の動き情報を取得した。加速度センサの出力は Bluetooth を使用して送信し、受信側のプログラムでテキストファイルに保存した。

(3) モーションキャプチャデータ

被験者にはモーションキャプチャ用のマーカを取り付けた帽子を装着してもらい、この帽子を 1 つのオブジェクトとして認識し、その x, y, z 座標と回転角度を計測した (図 4)。また、マーカを肩、肘、手首に取り付けた服を着てもらい、さらに中指にもマーカを巻き付けてもらうことにより、左右の 4 つずつのマーカのデータも収集した。



図 4 リジッドボディ

(4) 顔映像

テーブルの中心に各被験者の正面顔が撮影できる角度にウェブカメラ (Logicool : HD Pro Webcam C920t) を設置し、被験者の顔映像を取得した。被験者とウェブカメラの距離は約 1m であり、録画の際には被験者の顔をズームして録画した。顔映像の録画には、使用したウェブカメラが標準で搭載しているソフトウェアを使用した。

(5) 骨格情報

各被験者の背後に配置した Kinect センサから、対面にいる被験者の骨格情報と深度情報を取得した (図 5 参照)。

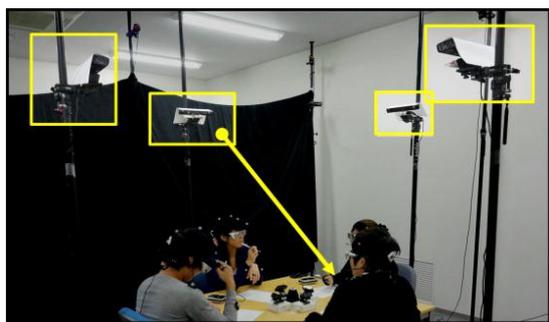


図 5 Kinect の配置

(6) 視線情報

被験者には視線追跡用の装置であるグラス型アイトラッカ (Tobii : Tobii Glasses Eye Tracker) を装着してもらい、各被験者の視線情報を取得した。視線データは SD カードに保存され、実験後に専用ソフトウェアの TobiiStudio で処理することにより、被験者視点のシーン映像と、これに重畳された被験者の注視点を avi 形式の映像データとして保存した。

(7) 性格特性

実験開始前に、NEO-FFI 性格特性評価テストを実施した。この性格テストは、BigFiveScale に基づき設定された 60 項目からなる質問紙形式のテストである。“活気のある所にいるのが好きだ”や、“抽象的な考え方や理論を楽しむことがよくある”といった質問項目に対し、“非常にそうだ”から“全くそうでない”の 5 段階で回答してもらうことにより、BigFive 性格特性項目である“経験への開放性”、“勤勉性”、“外向性”、“協調性”、“情緒不安定性”の 5 項目についての評価が得られる。

4. 議論状態把握モデル

音声と加速度センサのデータを用いて議論状態を把握するモデルを作成する。まず、各データの加工方法を以下に述べる。

4.1 音声情報

収録した音声データを 0.01 秒単位に分割し、各区間について、発話の有無と韻律情報を付与した。発話の有無については、各区間の各被験者の発話状態を 0, 1 で表現した。具体的には、音声分析ソフトウェア Praat により算出される音圧の値が一定以上の場合に 1 を、音圧の値が閾値以下である場合には 0 を与える。4 人の被験者に対して同様の処理を行うことにより、0.01 秒単位でのグループ構成員の発話状態を 0, 1 のビットパターンで表すことができる。0000 は誰も発話をしていない状態 (Ptn0)、0010 や 0001 のように、1 か所だけ値が 1 になっているパターンは一人の被験者のみが発話している状態 (Ptn1)、0110 や 0111 のように 2 か所以上が 1 であるパターンは複数の被験者が同時に発話している状態 (Ptn2) を表す。また、発話状態の変化も考慮するため、誰も発話していない状態への移行 (ShiftToPtn0)、一人の被験者のみが発話する状態への移行 (ShiftToPtn1) も含め、合計 5 種類の発話状態を定義した。

韻律情報については、0.01 秒の各区間における音圧を Praat により計測し、その値を使用した。

4.2 加速度情報

加速度センサは x, y, z の 3 軸における加速度、角速度を 33msec ごとに出力する。本研究では加速度センサから得られる加速度の情報を用いて、各被験者の頭部がその人の平均的な姿勢からどの程度変化したかを求め、それを被験者における頭部位置変化量として、以下の計算式により算出した。

$$t \text{ における変化量} = \sqrt{(x_t - \bar{x})^2 + (y_t - \bar{y})^2 + (z_t - \bar{z})^2}$$

ここで x_t, y_t, z_t は、任意の時点 t における x, y, z 軸の加速度を、 $\bar{x}, \bar{y}, \bar{z}$ はセッションを通した x, y, z 軸における加速度の平均を表す。

4.3 議論状態の分析

3 章で収集したディスカッションの様子をビデオにより観察したところ、次の 3 種類の議論状態に分類できると考えた。

- ① **ordered**: 参加者が順番に発言し、フロアマネジメントが整然と行われている状態
- ② **unordered**: 複数の人が同時に発言をし、発話がオーバーラップすることにより、フロアマネジメントが乱れている状態
- ③ **discontinuity**: だれも発言せず、話が途切れた状態

本研究では、これら 3 つの状態を自動的に判別できれば、議論が活発であるか停滞しているかを判断する有益な情報になると考え、これら 3 種類の議論状態を判別するモデルを作成する。

まず、教師データを作成するために、ビデオに対して、上記3種類の議論状態を、ビデオアノテーションツール ELAN を用いてアノテーションした。ここでは、実験で実施した3種類の課題のうち、提示された課題に対して解決案を話し合う、ケーススタディ型課題である“学園祭出店計画”のコーパスを分析対象とした。なお、音声と加速度のデータが一部欠損していたため、最終的には11グループ中8グループのデータを学習に用いた。

4.4 議論状態把握モデルの学習

4.1, 4.2 節で準備した 0.01 秒単位のデータを 1 秒単位で集計した頻度データを作成し、これを機械学習に用いた。具体的には、1 秒の区間において、4.1 節で定義した 5 種類のグループ発話状態の各発生回数、4 名の参加者の音圧の合計、および頭部位置変化量の合計を算出した。これら 7 種類の特徴量について、1 秒前、および 2 秒前の履歴情報も追加し、計 21 の特徴量を設定した。

不要な属性を削除するために、ラッパーアプローチを用いて、属性選択を行った。その結果、Ptn0, Ptn1, Ptn2 の頻度、音圧の合計、一秒前の区間における Ptn1, Ptn2 の頻度が残り、加速度の情報は選択されなかった。これら 6 種類の特徴量からなる学習データを用いて、SVM による機械学習を行った。カーネルには、4 次の多項式カーネルを使用し、C パラメータは 1.0 に設定した。

モデルの評価には leave-one-group-out 法を用いた。モデルの評価結果を表 1 に示す。議論の途切れである discontinuity とターン交代が整然、かつ円滑に行われている ordered の状態は、それぞれ F-measure で 0.755, 0.851 と高い精度が得られた。それに対し、フロアマネジメントが乱れている unordered の状態の推定は 0.467 にとどまった。その原因は再現率の低さにある。本モデルでは、unordered である箇所を十分な精度で検出できていないと言える。

表 1 各クラスに対する分類の評価結果

Class	Precision	Recall	F-Measure
discontinuity	0.807	0.709	0.755
ordered	0.798	0.911	0.851
unordered	0.742	0.341	0.467

4.5 議論状態把握モデルの考察

4.4 節で得られた議論状態推定モデルを学習に利用していないデータに適用した結果を図 6 に示す。グラフには、20 分間の議論における、discontinuity, ordered, unordered の各状態が推定された回数を 5 分単位で集計した結果を示している。

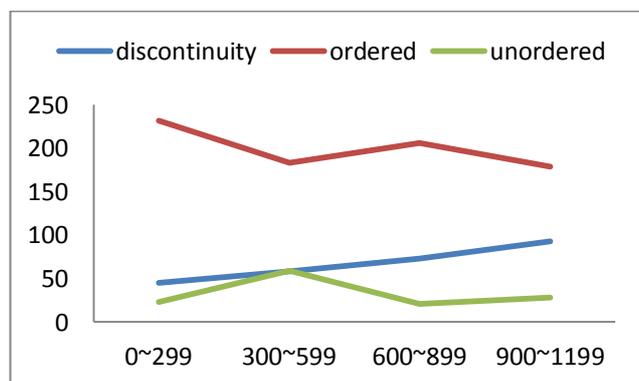


図 6 議論状態の推定

誰も発言しておらず、話が途切れた状態である discontinuity が多いと、議論が停滞している可能性が高いと考えられるが、この議論においては後半ほど discontinuity の推定回数が増加しており、時間経過に伴い議論が停滞していく状況が示唆される。実際のビデオを観察してもこの図が示すような傾向が見られることを確認した。

5. おわりに

本研究では、まず、グループディスカッションの対話収集実験を行うことにより、対話コーパスを構築した。このコーパスは 4 人で構成された被験者グループに対して、グループディスカッションの課題を与え、課題遂行中の各被験者の発話音声、頭部の動き量、モーションキャプチャデータ、顔向き映像、骨格情報、視線情報を計測することにより作成された、これらの行動データに加え、全被験者に質問紙による性格特性検査を行った。

次に、構築されたコーパスを観察し、discontinuity, ordered, unordered の 3 つの議論状態を定義し、発話状態のパターンと頭部位置変化量を学習データとし、この 3 状態を判別するモデルを SVM により作成した。その結果、発話状態のパターンのみを用いたモデルが得られた。さらに、作成したモデルを新たな対話データに適用し、議論状態推定結果から、議論が活発か、停滞しているかを判別できることを確認した。

本研究では、ビデオ観察に基づき議論状態を独自に定義したが、社会学や議論学の知見を利用して、議論状態の定義を再検討する必要があると考える。また、提案モデルは言語情報に依存しないため、今回対象とした議論以外にも適用できる可能性がある。今後、モデルの汎用性についても検証していく予定である。

参考文献

- [堀 2004] 堀公俊：ファシリテーション入門，日経文庫 (2004)。
- [Jovanovic 2005] N. Jovanovic, R. op den Akke, and A. Nijholt: A corpus for studying addressing behavior in multi-party dialogues, *In Proc. of The 6th SigDial conference on Discourse and Dialogue* (2005)。
- [Jurafsky 1997] D. Jurafsky, et al.: Automatic Detection of Discourse Structure for Speech Recognition and Understanding, *In Proc. of IEEE Workshop on Speech Recognition and Understanding* (1997)。
- [大本 2010] 大本義正, 戸田泰史, 植田一博, 西田豊明: 議論への参加態度と非言語情報に基づくファシリテーションの分析, *情報処理学会論文誌*, Vol.52, No.12, pp.1234–1245 (2011)。
- [Pianesi 2007] F. Pianesi, M. Zancanaro, B. Lepri, and A. Cappelletti: A multimodal annotated corpus of consensus decision making meetings, *Language Resources and Evaluation*, 41:409–429 (2007)。