

# 議論中の言語・非言語情報に基づく発散的／収束的発話の識別

## Classifying Divergent/Convergent Utterances in Group Discussions based on Verbal and Nonverbal Information

富山 健\*<sup>1</sup>  
Ken Tomiyama

高瀬 裕\*<sup>2</sup>  
Yutaka Takase

中野 有紀子\*<sup>2</sup>  
Yukiko Nakano

\*<sup>1</sup> 成蹊大学理工学部情報科学科  
Graduate School of Science and Technology, Seikei University

\*<sup>2</sup> 成蹊大学理工学部  
Faculty of Science and Technology, Seikei University

Aiming at developing a system that facilitates group discussions, this research proposes a model that classifies whether an utterance diverges the discussion or converges them. As prediction features, we used language information such as the number of nouns that appeared in the conversation for the first time and those already appeared in the preceding conversation, and non-verbal information such as prosodic information and utterance length were used. We created a three-class classification model: divergence, convergence, and others, using SVM. The accuracy of the model was 53.9% in leave-one-person-out cross validation.

### 1. はじめに

議論は意思決定やアイデア創出等、様々な目的で行われるが、より良い議論を行うためには、議論の整理役となり、合意形成や相互理解に向けて議論を進めていくファシリテータの存在が重要である。また、従来研究[堀 2009]から議論ファシリテーションには議論の進行を把握する必要があることがわかっており、そのためには、例えば、議論が今発散状態であるのか、収束状態であるかを認識することが有用である。発散とは議論を広げること、収束とは議論をまとめていくことを指す。ファシリテータはこの発散と収束をとらえながら適切なファシリテーション行動を選択・実行している。従って、本研究が目指す議論ファシリテーション支援システムを構築するためには議論中の発散と収束を識別することは重要な課題である。

そこで本研究では、言語・非言語情報から議論参加者の各発話が発散的であるか収束的であるかを識別するモデルの作成を目的とする。

### 2. 議論コーパス

本研究で使用する議論コーパスについて述べる。

1グループ3人×10グループ、男女30人の実験参加者が議論を行った。課題は外国人へ向けた一日観光コースをテーマに30分間話し合い、その後5分間でアピール点を決定してもらうというものである。最終的に「観光客の国籍と観光コースのキャッチフレーズ」「観光コースの詳細」「アピール点」を記入した用紙を提出させた。議論は「対象の外国人の国籍」「コースのキャッチフレーズ」「観光コースの作成」の順に行うよう指示した。議論前に参加者には議論終了後に作成した観光コースなどについて「観光スポットの数が多し」などの評価点によって評価されることが伝えられた。実際に行われた会話の例を図1に示す。この会話例は対象の外国人の国籍を決定した場面である。

次に図2に実験環境を示す。Kinectによる各参加者の頭部追跡データ、頭部に装着したモーションセンサによる20ms毎の加速度・角速度・地磁気データ、ヘッドセットマイクによる各参加者の音声データ、マイクロホンアレイによる議論全体の音声データ、ビデオカメラによる議論全体の動画が収集された。また

録画した動画から議論の書き起こしを作成した。実験は外部の音が入りにくい密室で行われ、被験者が議論を行う場を黒いカーテンと白い壁で囲い、議論参加者外部の影響を少なくするよう配慮した。

- |   |  |
|---|--|
| A | 「例えば、国名を絞ればいいので、欧米人にするか。」                                  |
| C | 「んー。中国圏ですよ。」   |
| A | 「中国人とかそういう意見にするか。」   |
| B | 「うん、うん、考えやすいのってどっちでしょうかね。」                                 |
| A | 「そうですね、どなたかこだわりがあれば。」                                      |
| B | 「えへへ、こだわりはある。」   |
| C | 「はは。まあでも、中国圏のほうがやっぱり想像がしやすいですよ、今ニュースとかでもよくやっているんで。うん。」     |
| B | 「ああ。」  |
| A | 「そうですね。伝統、文化とかは多分別に誰でも通用するので、買い物とかにもそこに時間をかけると、中国人っていうのは。」 |
| C | 「うん。」  |
| A | 「東京モダンとか、流行とかは、中国人であろうと、欧米の方であろうと別に関係ないことだと思うので。」          |
| C | 「まあ、そうですね。うん。中国人でいきましょうよ。」                                 |
| A | 「中国人で。」  |
| B | 「年齢とかはどうします。」  |

図1 会話例(対象の外国人の決定)

連絡先: 富山健, 成蹊大学理工学部研究科知的インタフェース研究室, 東京都武蔵野市吉祥寺北町3-3-1

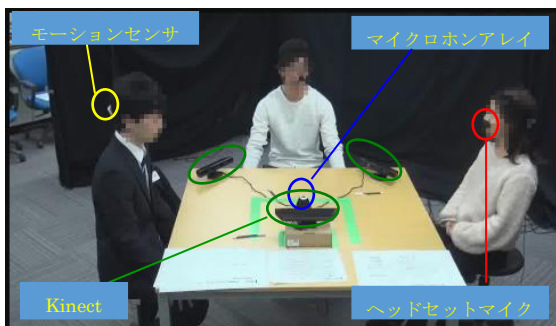


図 2 実験環境

### 3. 発話のアノテーション

発散/収束の発話を検出するモデルを機械学習により作成するために、教師データとなる発話カテゴリのアノテーション作業を行った。

録画した実験映像を観察したところ、議論中には発散的発話と収束的発話のほかに、同意を示す相槌的な発話や、共通認識を確立するための確認的な発話などが見られた。そこで、発話アノテーションでは、発散的発話・収束的発話・相槌・確認・その他の 5 つの発話カテゴリを設定した。これらの発話カテゴリの定義と具体例を表 1 に示す。定義の作成には[大本 2011]のファミリータの行動の定義や[堀 2009]を参考にした。

表 1 タグの定義と例

発散的発話	評価や選択を行わず、あらゆるアイデアを徹底的に出し切ること 例)「東京タワーに行くのはどう?」
収束的発話	効果性や実現性などいろんな観点でアイデアを評価したり、取捨選択したりして、絞込みをかけていくこと 例)「プラン 1 は定番スポットたくさんあるし、いいんじゃないか?」
相槌	話し手から送られた情報を聞き手が共有したことを伝えるいくつかの固定的な表現 例)「うん」「はい」
基盤化	議論参加者間が共通の認識を持つようになることに寄与する発話 例)「つまり、あなたが言いたいことは渋谷に行きたいということですよね?」
その他	上記以外の発話

自動で検出された音声区間に対して、発話カテゴリを手で付与した。全 9 グループをそれぞれ 3 分割し、約 10 分のファイルを 27 ファイル作成し、合計 11 人で各ファイルにつき 3 名のアノテータが作業を行った。その結果、全 15681 個の音声区間にたいして発話カテゴリが付与された。

その後、同一発話者の連続した 2 つの音声区間の間のポーズが 1 秒以内のものを統合する作業を行った。本研究ではこの作業の結果得られた各区間を 1 発話とした。その結果総発話数は 7052 個となった。また音声自動検出プログラムが誤って発話音声ではないものを検出してしまう場合があった。このような音声区間は分析から除外した。

各発話について 3 名がタグを付与したため、3 名中 2 名以上が選択したカテゴリをその発話のタグとした。その結果 3 名の付与したタグが一致した発話は全体の約 52%、2 名以上のアノテ

ータの判断が一致していた割合は約 95%であった。しかし、3 名の判断が全て異なっていた場合やタグが 1 つに決定しなかった場合には、その発話は分析対象からは除外し、その結果、学習に使用するデータ数は 6139 発話となった。

### 4. 機械学習の特徴量

発散・収束を判別するモデルの作成にあたって、使用する特徴量を下に記す。これらは発話単位で算出される。[仙石 2016]では意見内容の発散/収束判定を名詞の新規性により判定している。これを参考に特徴量として新規/既出名詞の数を加えた。また導出方法には[岡田 2016]を参考にした。

#### (1) 新規/既出名詞の数

発話の書き起こしを形態素解析器 Mecab にかき、名詞を抽出する。抽出した名詞がそれまでの議論で出現した名詞と一致すれば既出名詞、一致しなければ新規名詞とし、それぞれの個数を数え上げる。一度出現した名詞は既出名詞一覧に登録され、以後の新規/既出名詞の判断に用いられる。

#### (2) 現発話と直前の発話間の一致/不一致名詞数

直前の発話と現発話を形態素解析にかき、名詞を抽出し、両発話に共通する名詞の数を共通名詞数とする。また、現発話に出現し、直前の発話には出現しない名詞の個数を不一致名詞数とする。

#### (3) 発話長(時間長と形態素数)

発話長を示す特徴量として、発話区間の時間長 ms と発話を構成する形態素数の 2 種類を使用した。発話者の話速により、同じ時間長でも形態素数が異なるため、発話の形態素数も発話長として特徴量とした。

#### (4) オーバーラップ時間

2 つの連続する発話のオーバーラップ時間 ms を特徴量とする。1 人の発話に 2 人以上の発話がオーバーラップしている場合は、それらの合計を算出する。下の図 3 と図 4 を用いて具体的に説明する。図 3 と図 4 は発話者 3 の発話が発話者 1 と発話者 2 の発話にオーバーラップしている場合の図である。図 3 より発話者 1 と発話者 3 のオーバーラップ時間を A、図 4 より発話者 2 と発話者 3 のオーバーラップ時間を B とすると発話者 3 のオーバーラップ時間はこれらを合計した A+B で計算される。

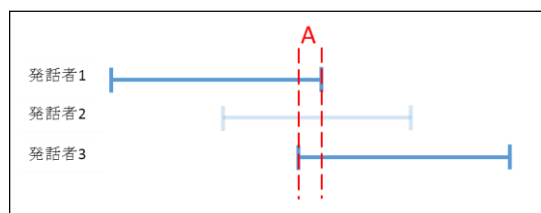


図 3 発話者 1 と発話者 3 のオーバーラップ時間

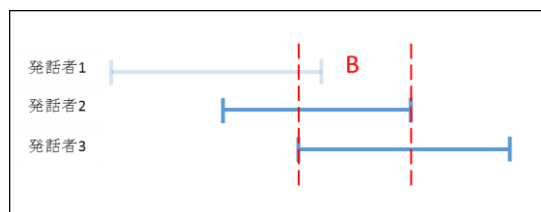


図 4 発話者 2 と発話者 3 のオーバーラップ時間

## (5) 頭部回転量

発話が行われている時間を対象に、20 ms の時間窓を設定し、各区間における頭部のヨー角回転量(度)をデータとする。各区間における最大値・最小値・平均値・分散値をそれぞれ特徴量とする。

## (6) 音声インテンシティ

音声解析ソフト Praat を用いて、収録した発話音声のインテンシティ(dB)を計測する。各発話区間において 10ms の時間窓を設定し、その区間におけるインテンシティの最大値・最小値・平均値・分散値をそれぞれ特徴量とする。

またこれらの各特徴量について、発散的発話と収束的発話の間で F 検定を行った後、等分散である場合、または等分散でない場合、それぞれに対応する t 検定を行った。その結果、インテンシティの平均値と頭部ヨー角回転量の平均値と分散値において、発散と収束の間で差が有意ではなかった。そのため、これらの 3 つの特徴量は予測に有用でないと考え、これらを除外し、合計 12 個の特徴量を用いてモデルの作成を行った。

## 5. 識別モデルの作成と評価

発散的発話・収束的発話・その他の 3 値分類モデルを作成する。モデル作成においては、4 章で定義・選定した特徴量を使用し、決定木学習、RandomForest, NaiveBayes, SVM(線形カーネル, C パラメータ 1.0)の 4 つの学習手法を用いた。使用した学習データは各 27 名のデータセットごとに 3 値の中で最も少なかったタグでサンプリングを行ったデータセットである。leave-one-person-out によるモデルの評価を行った結果を表 2 示す。SVM モデルの精度(0.539)が最も高く、チャンスレベルの 33%をベースラインにすれば、約 20%高い精度が得られた。また最も精度の高かった SVM モデルについて適合率、再現率、F 値を表 3 に示す。収束的発話よりも発散的発話のほうが高い評価であった。これは、収束的発話を識別するのに有用であると考えていた発話内既存名詞がうまく機能しなかったためと考える。

表 2 3 値分類モデルの精度

	正答率
決定木	0.496
RandomForest	0.533
NaiveBayes	0.492
SVM	0.539

表 3 SVM(線形カーネル, C パラメータ 1.0)モデルの評価

	適合率	再現率	F 値
発散的発話	0.642	0.526	0.578
収束的発話	0.448	0.359	0.398
その他	0.532	0.733	0.616

さらに、SVM モデルにおける各特徴量の重みを調べた。その結果を表 4 に示す。発話内新規名詞数の値が最も高く、発散的発話と収束的発話を識別する能力が高いと考えられる。また、直前の発話と一致する名詞数、直前の発話と一致しない名詞数、オーバーラップ時間、インテンシティの最大値・最小値、頭部ヨー角回転量の最小値、発話長(形態素数)も有用であると考えられる。

表 4 各特徴量の重み

特徴量名	重み
発話内新規名詞数	-7.483
発話内既存名詞数	0.147
直前の発話と一致する名詞数	2.464
直前の発話と一致しない名詞数	-2.998
発話長(時間)	-0.540
オーバーラップ時間	3.383
インテンシティの最大値	-1.127
インテンシティの最小値	2.299
インテンシティの分散値	1.144
頭部ヨー角回転量の最大値	0.474
頭部ヨー角回転量の最小値	1.195
発話長(形態素数)	-3.891

## 6. まとめと今後の課題

本研究では、議論中の発話が発散的か収束的であるかを識別するためのモデルを決定木、RandomForest, NaiveBayes, SVM の手法を用いて作成した。leave-one-person-out 法による評価の結果、SVM モデルで分類精度 53.9%の性能を得た。

今後の課題として、モデルの精度を上げるため、特徴量の追加やアノテーション手法の改善などがあげられる。特徴量の追加については、音声のピッチ、視線情報などの追加が考えられる。また、教師データとなるアノテーションにおいて、3 名のアノテータ全員の判断が一致するものはまだ約半数にとどまっており、アノテータの追加、定義の見直しなどを行うことにより、一致率を向上させることが必要である。

## 謝辞

本実験は KDDI 研究所(現 KDDI 総合研究所)と共同で実施したものである。

## 参考文献

- [大本 2011] 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 3659-3670, 2011.
- [仙石 2016] 仙石 晃久, 伊藤 孝行, 藤田 桂英, 白松 俊: Web 上での大規模議論における議論ツリーによる意見集約支援, 第 30 回人工知能大会, 3F4-2, 2016.
- [岡田 2016] 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: グループディスカッションにおいて現れる コミュニケーション能力のマルチモーダル分析, 人工知能学会論文誌, AI30-E, 2016.
- [堀 2009] 堀 公俊, 加藤 彰: ロジカル・ディスカッション チーム思考の整理術, 日本経済新聞出版社, 2009.