

# 説得力および納得度の推定に向けた Web デイベートデータセットの構築

大杉康仁<sup>1</sup> 吉田明弘<sup>1</sup> 中辻真<sup>2</sup>

<sup>1</sup>株式会社 NTT ドコモ <sup>2</sup>NTT 人間情報研究所

yasuhito.oosugi.fg@nttdocomo.com

## 概要

話者の説得力、および、聴者の納得度を自動で推定することが可能となれば、自学自習による練習機会が増え、営業スキルやコミュニケーションスキルの向上につながる。本研究では、オンライン会議において話者の説得力や聴者の納得度をラベル付けしたデータセットを構築した。話者と聴者が明示的に交代するディベート形式を採用し、肯定側 1 人・否定側 1 人・判定者 1 の合計 3 人での会議を 115 セッション収録した。第三者から見た説得力や論理性、聴者の納得度などのアノテーションを実施した。ラベルの分析からは、説得力と論理性には相関係数 0.796 の強い正の相関が、説得力と判定者の納得度については相関係数 0.554 の正の相関が、それぞれ見られた。

## 1 はじめに

商談の成功率を高める上で営業担当者の説得力は重要である。人どうしで営業の練習を行いフィードバックをすることで説得力などのコミュニケーションスキルを高めることができるが、訓練コストが高いため練習回数を増やすことが困難である。

そこで、説得力をはじめとした営業スキルやコミュニケーションスキルを自動推定するツールへの期待が高まっている。コミュニケーションスキルの分析や自動推定ツールの構築のためにはコーパスやデータセットが不可欠である。従来より映画へのレビューを録画した動画に対する話者の説得力をスコア化した POM データセット [1] や、グループディスカッション中の話者のコミュニケーションスキルを評価した MATRICS コーパス [2]、遠隔操作アンドロイドが人間を説得する対話を収集したコーパス [3] など様々なコーパスやデータセットが提案されている。聴者の納得度に着目したデータセットとして

は、オンラインでのカウンセリングを受けた被験者自身が納得度を評価したコーパス [4] がある。しかし、話者の説得力と聴者の納得度を同時に評価したコーパスは少なく、説得力と納得度の関連性については調査の余地があると考えられる。

そこで本研究では、話者の説得力と聴者の納得度をアノテーションしたデータセットを構築した。近年急速に普及したオンライン会議ツールを利用し、話者の説得力が重要となるディベート形式の動画を 115 本収録した。収録した動画について第三者から見た話者の説得力、聴者の納得度をアノテーションし、それぞれのラベルについて分析を行った。

## 2 関連研究

話者の説得力に関するデータセットとして POM [1] があるが、これは一人の話者が映画に対するレビューを行う独話であり、次のパートで話者と聴者が交代するディベートである本研究とは異なる。本研究と同様にディベートを採用した日本語コーパスとして Kyutech デイベートコーパスがある [5]。これは、4 人の対面での会議を録画したものであり 2 人で 1 つの立場を形成している点で本研究と異なる。

英語テキストに閉じたディベートにおいて、どちらの立場が説得力のある主張をしていたかを大規模言語モデルを用いて推定する手法も提案されている [6]。GPT-4 を用いた手法では人間と同程度の精度を達成したと報告されているが、人間の精度も 60% 程度と高くはないためタスク自体の難易度が高い可能性も指摘されている。対面会議において、音声・動画・テキストを扱うマルチモーダルなアプローチで主張の品質を推定する手法 [7] や話者の説得力の高低を推定する手法 [8] が提案されている。しかし、本研究のようなオンライン会議においては、上半身の動きを撮影しにくく、また対面での会議と参加者

の振る舞いが異なる可能性もあり、推定モデルの特徴量や構造には検討の余地があると考える。

### 3 Web ディベートデータセット

#### 3.1 収集条件

**問題設定** 我々が提示するテーマに沿って、肯定側 1 人・否定側 1 人・判定者 1 人の 3 人でのディベートを収録した。テーマとしては、「朝ごはんはパンかお米のどちらがいいか？」などの参加者が回答しやすい内容を採用し、5 テーマを 1 セットとし 5 セット（合計 25 テーマ）を用意した<sup>1)</sup>。以降、ある 1 つのテーマに対するディベートを「1 セッション」と呼ぶ。1 セッションは肯定側/否定側の立論や反駁など、肯定側と否定側で議論を行う 8 つのパートと、判定者が勝者を決定する判定パートの合計 9 パートに分けられる<sup>2)</sup>。判定者を司会進行役とし、基本的にこの流れに沿うよう収録を実施した。ディベートの勝者の判定においては、「立論で出した意見は重要か」「反対尋問・反駁で相手の意見を崩せたか」などを 3 段階で評価しその合計スコアで勝敗を決定させた。

**ツール** オンライン会議ツール Zoom<sup>3)</sup> を利用し、合計で 115 セッション収録した。音声のオーバーラップを防ぐため、各話者を個別の会議室に配置し、PC に備え付けのマイク・カメラを利用し、各話者の顔が映るように設定を行った。1 セッションにつき、画面全体の動画 1 つと、各話者の個別音声 3 つが得られた。

**収録参加者** 参加者は、20 代から 50 代の男性 12 人と女性 29 人の合計 41 人であった。参加者が複数のセッションに参加することを許容した。ディベートの勝敗や説得力の有無と性格特性の相関を調査するため、各参加者に対し性格特性に関するアンケート [9] を実施した。また、各セッション終了後に参加者自身が思う説得力や納得度を問うアンケートも実施した。

**資料共有** オンライン会議の利点として、資料共有が簡単に実施できることが挙げられる。そこで、Wikipedia などの Web ページを共有資料として利用

1) テーマによっては肯定や否定の表現が適切でないケースがあるが、便宜上「肯定側」「否定側」と表現する。この例の場合「パンが良い」が「肯定側」、「お米が良い」が「否定側」となる。

2) <https://www.burningmind.jp/debate/>

3) <https://www.zoom.com/ja/products/virtual-meetings/>

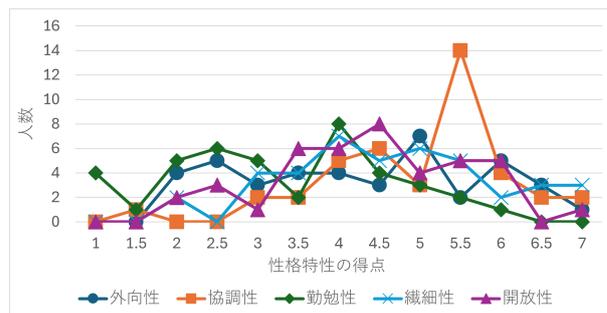


図 1 Big-five 性格特性の人数分布

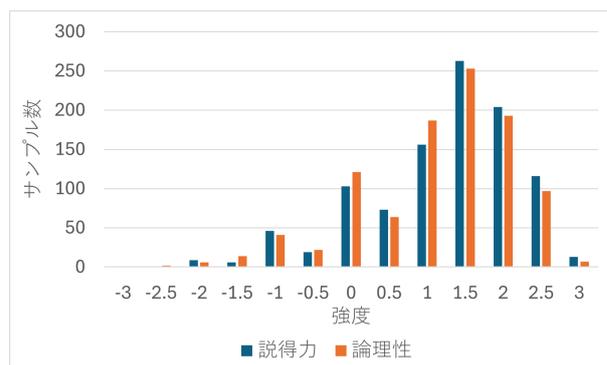


図 2 論理性と説得力の分布：数字が大きい方が論理性または説得力が高いことを意味する。

可能とした。ただし、参加者が利用可能な情報源を可能な限り統一するため、Web ページは筆者らが事前に収集したものを利用するよう指示を出した。

#### 3.2 収集結果

**収録参加者の性格特性** 図 1 に示すように、協調性のスコアが高い参加者が多く、他の性格特性については比較的分布はなだらかであった。

**動画分割** ディベートの流れに沿って動画を分割するために、音声認識結果に基づいて自動的にパート境界を抽出し、人手で確認することで各動画を 9 つのパートに分割した。なお、ディベートの規定の流れから外れた場合については、パートの順番の変更や雑談の除去について目視で対応したが、一部のデータについてパート境界に不備があったため、それらを除外した。結果として、合計時間 39.6 時間、平均 141.4 秒、最大 645.8 秒、最小 29.6 秒の、1008 個のパート動画が得られた。

**資料共有の割合** 資料の画面共有の割合を表 1 に示す。全体としては 22.4% の動画に対し画面共有がなされていた。ディベートの各パートにおいては、肯定側と否定側の双方で立論時の画面共有割合が高く、自身の主張の補強として資料が活用されたとと言える。

表1 資料の画面共有の割合

主な話者	パート項目	データ数	割合 [%]
肯定側	立論	111	61.3
否定側	反対尋問	112	13.4
否定側	立論	111	45.9
肯定側	反対尋問	111	14.4
否定側	反駁	115	20.9
肯定側	反駁	109	22.9
否定側	最終弁論	114	7.0
肯定側	最終弁論	112	11.6
判定者	判定	113	5.3
	合計	1008	22.4

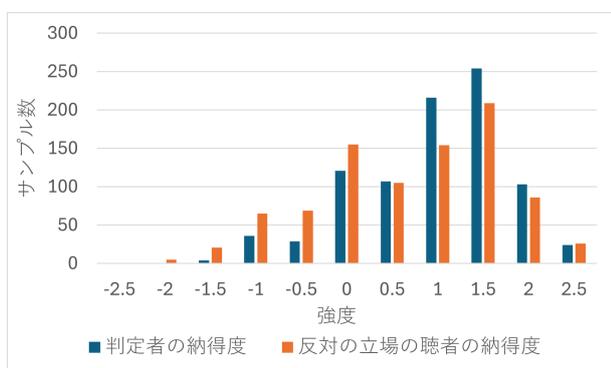


図3 判定者または反対の立場の聴者の納得度の分布：数字が大きい方が納得度が高いことを意味する。

### 3.3 アノテーション

前節で得られた各パートの動画について、主な話者に対する第三者から見た説得力や話の論理的展開（論理性）、および主な聴者に対する第三者から見た納得度をアノテーションした。表2に示すように各パートによってアノテーション項目を変更した。最終弁論までのパートについては、主な話者の論理性・説得力と主な聴者の納得度をそれぞれ7段階のリッカート尺度で評価した。判定パートについては、主な話者（判定者）の論理性・説得力に加え、判定に対する肯定側と否定側の納得度をそれぞれ7段階のリッカート尺度で評価した。各項目について、2人の異なるアノテータが評価を行った。アノテータ間の一致度を調査するため、アノテータ間、および、各アノテータと平均値との間、の2通りのクリップドルフの $\alpha$ 係数を算出した。前者は0.3程度と一致度が低い結果となったが、後者については0.7以上であり、先行研究[1, 8]と同様に中程度の信頼度と言えるため、本研究では平均値をそのラベルの真値として扱う。

## 4 分析

### 4.1 説得力と論理性の関連性

話者の論理性と説得力のラベル分布を図2に示す。これらの分布は似通っており、相関係数は0.796と強い正の相関が見られた。論理性は主に発話内容（言語モダル）に関連すると考えられるため、話者の説得力推定にはテキスト情報が有用である可能性が示唆された。これは、伊藤らの説得力推定モデル[8]において、unimodalモデルでは言語モダルの精度が最も高かったことを補強すると考えられる。

### 4.2 最終弁論までのパートにおける説得力と納得度の関連性

**納得度の分布** 判定者または反対の立場の聴者の納得度ラベル分布を図3に示す。どちらの分布も納得度が高い方向に偏っていたが、判定者の納得度分布はスコア1.5に山を持つ単峰性分布と言える。これは、判定者は肯定側と否定側のどちらの意見も等しく聞き勝敗を決める必要があるため、どちらの主張も受け止め納得していると第三者から評価された可能性がある。一方で、反対の立場の聴者の納得度はスコア1.5と0に山を持つ二峰性分布と言える。また、スコアが0未満のサンプルも多かった。これは、聴者自身が次のパートで反駁や反対意見を述べる必要があるため、判定者よりも納得度の強弱が表れた可能性がある。

**話者の説得力との相関** 話者の説得力と、判定者の納得度または反対の立場の聴者の納得度の相関係数はそれぞれ0.554と0.433であり中程度の正の相関と言える。これは、同じ話者・発話内容・発話スタイルであっても納得する人とならない人がいるが、説得力が高ければ納得する人の割合が多くなるという直観に合致すると言える。また、聴者の納得度を推定するためには、聴者自身の表情やしぐさなどの情報以外にも、話者の情報、特に説得力に起因する情報が有用である可能性が示唆された。ディベートにおける立場で比較すると反対の立場の聴者の方が相関係数が小さく、前節の分布の違いが表れていると考えられる。

### 4.3 判定パートにおける説得力と納得度の関連性

判定パートにおける、肯定側または否定側の納得度の分布を図4に示す。立場によらず、スコア1.5

表2 説得力・納得度のラベルとクリッペンドルフの $\alpha$ 係数

	最終弁論まで	判定	データ数	平均値との $\alpha$	アノテーション間の $\alpha$
話者の論理性	✓	✓	1008	0.795	0.320
話者の説得力	✓	✓	1008	0.792	0.315
判定者の納得度	✓		895	0.817	0.381
反対の立場の聴者の納得度	✓		895	0.794	0.322
判定に対する肯定側の納得度		✓	113	0.799	0.341
判定に対する否定側の納得度		✓	113	0.801	0.352

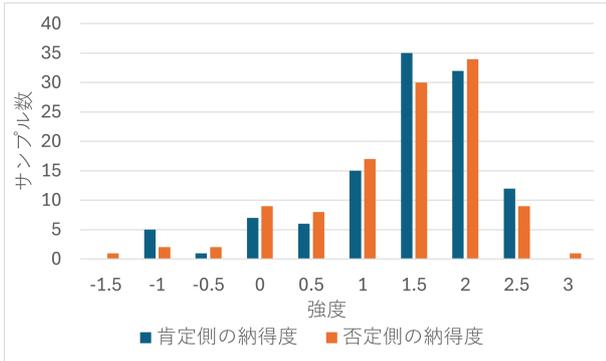


図4 判定に対する肯定側または否定側の納得度の分布：数字が大きい方が納得度が高いことを意味する。

もしくは2と納得度の高い位置に山を持つ分布であった。また、話者（判定者）の説得力と肯定側または否定側の納得度の相関係数は、それぞれ0.364と0.326であり、最終弁論までの説得力と納得度の相関よりも低くなった。判定者の説得力は高いが、判定結果に自身の意見が反映されなかったため納得していないケースがある可能性があり、今後も調査を継続する。

## 5 まとめと今後の課題

本研究では、話者の説得力および聴者の納得度の自動推定に向けて、オンライン会議におけるディベートを収録したデータセットを構築した。ラベルの分析を通じて、話者の説得力と論理性には強い正の相関があったため、説得力の推定には言語モデルが有用である可能性が示唆された。また、聴者の納得度と話者の説得力にも正の相関があるため、聴者の納得度の推定には聴者の表情などのマルチモーダルな情報以外にも、話者の発話内容などの言語情報が有用であると考えられる。

本稿では第三者から見た納得度に着目して分析を行ったが、ディベート参加者自身の納得度に関しては未調査である。納得度は説得力よりも参加者の内面に根ざした指標と考えられるため、第三者から

見た度合と本人の度合が異なる可能性があり、今後も調査を続ける。また、性格特性やディベートの勝敗、共有資料の内容などと、説得力および納得度の関連性についても調査する。さらに、話者の説得力や聴者の納得度を自動で推定する手法、特に大規模言語モデルを用いた手法についても今後検討していく。

## 参考文献

- [1] Sunghyun Park, Han Suk Shim, Moitrey Chatterjee, Kenji Sagae, and Louis-Philippe Morency. Computational analysis of persuasiveness in social multimedia: A novel dataset and multimodal prediction approach. In **Proceedings of the 16th International Conference on Multimodal Interaction**, p. 50–57, New York, NY, USA, 2014. Association for Computing Machinery.
- [2] 佑樹林, 美巳雄二瓶, 有紀子中野, 宏軒黄, 将吾岡田. 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, pp. 1217–1227, apr 2015.
- [3] 有岡無敵, 山本賢太, 井上昂治, 河原達也, 中村哲, 吉野幸一郎. 遠隔操作アンドロイドを用いたマルチモーダル説得対話コーパスの収集と分析. 言語処理学会 第28回年次大会 発表論文集, pp. 185–190, 2022.
- [4] 松隈亮太, 岡田将吾, 松本妹子, 中元淳. オンラインカウンセリング対話データコーパスの構築と動作シンクローニーの分析. 人工知能学会 全国大会論文集, pp. 216OS9b03–216OS9b03, 2022.
- [5] Tsukasa Shiota and Kazutaka Shimada. Annotation and multi-modal methods for quality assessment of multi-party discussion. In **Proceedings of the 36th Pacific Asia Conference on Language, Information and Computation**, pp. 175–182. Association for Computational Linguistics, October 2022.
- [6] Paula Rescala, Manoel Horta Ribeiro, Tiancheng Hu, and Robert West. Can language models recognize convincing arguments? In Yaser Al-Onaizan, Mohit Bansal, and Yun-Nung Chen, editors, **Findings of the Association for Computational Linguistics: EMNLP 2024**, pp. 8826–8837. Association for Computational Linguistics, November 2024.
- [7] Md Kamrul Hasan, James Spann, Masum Hasan, Md Sai-ful Islam, Kurtis Haut, Rada Mihalcea, and Ehsan Hoque. Hitting your MARQ: Multimodal ARgument quality as-

essment in long debate video. In **Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing**, pp. 6387–6397, November 2021.

- [8] 伊藤温志, 坂戸達陽, 中野有紀子, 二瓶芙巳雄, 石井亮, 深山篤, 中村高雄. グループディスカッションにおける説得力推定のためのマルチパーティモデル. 人工知能学会 全国大会論文集, pp. 3H3OS12a02–3H3OS12a02, 2022.
- [9] 小塩真司, 阿部晋吾, Pino Cutrone. 日本語版 ten item personality inventory (tipi-j) 作成の試み. パーソナリティ研究, Vol. 21, No. 1, pp. 40–52, 2012.