

明治大学大学院

2025年度 修士論文

論文題名 競技ディベートの試合特徴理解に向けた

反論構造の定量化と可視化

先端数理科学研究科 先端メディアサイエンス専攻

指導教員名 中村聡史

本人氏名 福井雅弘

2025年度 修士学位請求論文

競技ディベートの試合特徴理解に向けた
反論構造の定量化と可視化

明治大学大学院先端数理科学研究科
先端メディアサイエンス専攻

福井 雅弘

Master's Thesis

Quantifying and Visualizing Rebuttal Structures
towards the Understanding of the Characteristics of
Competitive Debate Rounds

Frontier Media Science Program,
Graduate School of Advanced Mathematical Sciences,
Meiji University

Masahiro Fukui

概要

即興型の競技ディベートでは、ディベーターが建設的な反論を通じて対話を深めることが重要である。しかし、複数の話題が同時並行で展開される競技の特性上、試合全体を俯瞰して議論が深まっているかを把握することは難しい。従来の競技ディベートのフィードバックには、個々の発言内容の良し悪しや特定の論題に特化した反論の仕方といった局所的で定性的なものが多く、試合全体を通じてどのように反論が交錯し、議論がどのように深まったかを構造的に捉える視点から議論を構築するスキルを学習することは困難である。そこで本研究では、反論の個別の内容ではなく試合全体の反論構造に着目し、ディベートの評価と学習を支援する手法について調査することを目的に、反論構造の定量的特徴を明らかにし、その知見に基づいて可視化システムを設計・実装するという二つの段階に分けて手法の提案や検証を行った。具体的には、(1) 議論の噛み合い度合いを定量的に評価する手法の構築と精度の検証、(2) 反論構造の可視化により他者の参考になる試合動画の探索を支援するシステム「DebaTube」の提案と利用実験による検証、(3) 自身が参加した試合の可視化による振り返りを支援するシステム「DebaTube Live」の実装と実際の練習試合での検証を行った。反論構造の定量化と可視化を活用することで、評価と学習の両面から競技ディベートの教育的価値を高め、より建設的で質の高い議論スキルの習得を支援することを目指す。

まず、反論構造の定量的な特徴から議論の噛み合い度合いを推定できるかを調査した。その結果、再反論のラリーの頻度、遠い反論の割合、同じ主張への反論の間隔という3つの特徴量が議論の噛み合い度合いと相関することが明らかになり、勝敗予測に依存しない議論評価の基盤が確立された。

次に、これらの特徴量を視覚的に把握できる可視化手法を設計した。具体的には、ラリーの深さを追跡できるよう肯定側と否定側のノードを左右に分離し、反論の遠さや順序を把握できるようノードを上から時系列順に配置し、同じ発言への反論の間隔を直感的に理解できるよう各ノードを発言の長さによらず等幅で表現した。この可視化デザインに基づき、LLMを用いた反論構造の自動推定手法を開発し、他者の試合動画探索を支援するシステム「DebaTube」を実装した。利用実験の結果、ディベート経験者が反論構

造の可視化を手がかりに、望ましい構造的特徴を持つ試合の動画を効率的に探索できる可能性が示唆された。

最後に、自分が参加した試合の振り返りを支援するシステム「DebaTube Live」を実装した。本システムでは、ディベート特有のドメイン知識を反映したプロンプト設計と並列処理の導入により、約1時間の試合録音に対して約7分で反論構造を可視化できる実用的な処理速度を実現した。また、実際の練習試合を対象とした利用実験を実施した結果、提案システムを使用した群で反論構造の特徴に関する言及が顕著に増加し、主観評価でも議論の流れの整理や想起に関する項目で有意な改善が確認された。

以上の結果から、反論構造を計算的に扱うことで、ディベーターが試合の構造的特徴を俯瞰的に理解できることが示唆された。この効果により、その場の話題に左右されすぎず、議論の流れに沿った視野の広いディベートを展開するスキルの向上が期待できる。反論構造の定量化と可視化は、計算的手法とディベート学習支援を融合させ、ディベーターがより俯瞰的に議論を展開するスキルを高めるための基礎を築いた。本研究が、競技ディベートの教育的価値をさらに高め、世界中のディベーターのディベート学習に貢献することが期待される。

Abstract

In impromptu competitive debate, it is essential for debaters to deepen discussions interactively through constructive rebuttals. However, due to the nature of the competition where multiple topics are debated simultaneously, it is difficult to grasp the overall depth of argumentation while maintaining a broad perspective of view of the entire round. Traditional debate evaluation and learning often provide qualitative feedback on localized aspects, such as the quality of individual statements or topic-specific rebuttal techniques. There has been a lack of structural perspective for understanding how rebuttals intersect throughout a round and to what extent arguments have deepened. This research therefore focused on the overall rebuttal structure of rounds rather than the content of individual rebuttals, investigating methods to support debate evaluation and learning.

Our approach was phased: first clarifying quantitative features of rebuttal structures, then designing and implementing a visualization system based on these findings. Specifically, we conducted: (1) construction and validation of a method for quantitatively evaluating dialogue quality; (2) implementation and user study of “DebaTube”, a system that supports searching for instructive round videos through rebuttal structure visualization; and (3) implementation and validation of “DebaTube Live”, a system supporting reflection on one’s own rounds through visualization, tested in actual practice debates. By utilizing quantification and visualization of rebuttal structures, we aim to enhance the educational value of competitive debate from both evaluation and learning perspectives, supporting the acquisition of more constructive and higher-quality argumentation skills.

First, we investigated whether quantitative features of rebuttal structures reflect dialogue quality. Results revealed that three features including frequency of rebuttal rallies, proportion of distant rebuttals, and intervals between rebuttals to the same claim—correlate with dialogue quality, establishing a foundation for argument evaluation independent of win or loss prediction.

Next, we designed visualization methods that allow these features to be grasped visu-

ally. Specifically, we separated affirmative and negative side nodes left and right to track rally depth, arranged nodes chronologically from top to bottom to understand rebuttal distance and sequence, and represented each node with equal width regardless of statement length to enable intuitive understanding of intervals between rebuttals to the same statement. Based on this visualization design, we developed an automatic rebuttal structure estimation method using LLMs and implemented “DebaTube”, a system supporting the search for others’ round videos. User studies suggested that experienced debaters could efficiently search for videos of rounds with desirable structural features using the rebuttal structure visualization as a guide.

Finally, we implemented “DebaTube Live”, a system supporting reflection on rounds one has participated in. Through prompt design reflecting debate-specific domain knowledge and parallel processing, this system achieved practical processing speeds, visualizing rebuttal structures for approximately one hour round recordings in about seven minutes. User studies conducted with actual practice debates showed that the group using the proposed system exhibited a marked increase in references to rebuttal structure features, with subjective evaluations confirming significant improvements in items related to organizing and recalling the flow of arguments.

These results suggest that computational handling of rebuttal structures enables debaters to gain a broad perspective of view of structural round features and conduct effective reflection. This effect is expected to enhance skills for developing debates with broader perspective that follow the flow of arguments without being overly influenced by immediate topics. The quantification and visualization of rebuttal structures has established a foundation for integrating computational methods with debate learning support, helping debaters develop skills for approaching arguments from a broad perspective of view. We hope this research will further enhance the educational value of competitive debate and contribute to improving debaters’ argumentation skills.

目次

第1章	はじめに	1
1.1	競技ディベートの教育的意義と課題	1
1.2	反論構造の重要性	2
1.3	本研究の目的	3
第2章	関連研究	5
2.1	計算議論学と論述文における議論構造の解析	5
2.2	複数人間の対話的議論における議論の質の評価	6
2.3	反論の自動判定	8
2.4	テキスト提示による議論構造の理解支援	9
2.5	可視化による議論構造の理解支援	9
第3章	議論の噛み合い度合いの評価のための反論構造の定量化	12
3.1	反論構造のモデル化	13
3.1.1	用語の定義	13
3.1.2	反論の構造的特徴に基づく仮説	13
3.1.3	反論構造の特徴量の定義	14
3.1.4	反論構造の特徴量の解釈	16
3.2	コーパスの構築	17
3.2.1	データ構造	17
3.2.2	議論の噛み合い度合いの評価方法	18
3.3	結果	19
3.3.1	反論構造の特徴量と議論の噛み合い度合いの相関	19
3.3.2	特徴量の重要度の分析	20
3.4	分析と考察	20
3.4.1	評価者間信頼性	20
3.4.2	モデルの精度に関する詳細分析	21

3.5	小括	22
第4章	DebaTube の提案と評価	24
4.1	LLM を用いた反論構造の自動推定手法	25
4.2	反論構造の可視化のデザイン	27
4.3	DebaTube	28
4.3.1	必要要件	28
4.3.2	実装とシステム概要	30
4.4	予備調査	30
4.4.1	タスクの概要	31
4.4.2	タスク用システム	32
4.5	結果	32
4.6	分析と考察	33
4.6.1	動画探索行動の統計量	33
4.6.2	動画探索行動の時系列推移	35
4.6.3	試合動画の選択理由	36
4.7	小括	38
第5章	DebaTube Live の実装と利用実験による検証	39
5.1	LLM を用いた反論構造の自動推定手法の改善	40
5.2	DebaTube Live	41
5.2.1	必要要件	41
5.2.2	実装	42
5.2.3	利用方法	42
5.3	実験	43
5.4	British Parliamentary Style の概要	44
5.4.1	実験設計	45
5.4.2	実験システム	46
5.4.3	実験手順	46
5.4.4	振り返り内容の分析手順	47
5.5	結果	48
5.5.1	システムの動作	48

5.5.2	反論構造に着目した振り返りの傾向	49
5.5.3	振り返りのしやすさに関するアンケート結果	49
5.6	分析	51
5.6.1	反論構造への着目の促進	51
5.6.2	振り返りのしやすさの主観評価	52
5.7	考察	53
5.7.1	試合1：本議会は、学校は歴史を偉大な人物の産物としてではなく、 社会的および制度的要因の産物として教えるべきであると信じる	54
5.7.2	試合2：高齢化が進む社会において、本議会は、政府が国際的なマッ チング（結婚紹介）機関に補助金の拠出を支持する	54
5.7.3	試合3：実現可能性を考慮しなければ、本議会はウラジーミル・プー チンを暗殺する	56
5.7.4	小括	58
第6章	総合考察と展望	60
6.1	総合考察	60
6.2	限界と展望	61
第7章	結論	63
第8章	謝辞	65

第1章 はじめに

1.1 競技ディベートの教育的意義と課題

即興型ディベートとは、論題とよばれる賛否両論あるトピックに対して肯定側と否定側にランダムに分かれ、中立のジャッジを説得する知のスポーツである。この競技では、各チームの立場と持ち時間、話す順序が明確に定められており、限られた時間の中でジャッジを論理と感情の両輪で納得させる必要がある [1]。競技ディベートは、論理的思考力、表現能力、迅速な分析力、議論構築能力、反論技術など、参加者の総合的な能力を試す構造化された競技形式のコミュニケーションである [2,3]。

このディベートは、単なる勝敗を競うだけでなく [4]、参加者が批判的思考力や多角的な視点から建設的な議論を行うスキルを養うための教育的に重要な取り組みとしても位置付けられている [5]。過去のディベートを分析することで、参加者は効果的な議論の戦略を学び [6]、一般的な誤りを認識し [7]、優れた議論の立て方を理解することができる [8]。実際に、ディベートが批判的思考力の向上やコラボレーション能力、対話、コミュニケーション能力の改善に効果的であることが実証されている [9-11]。また、読解力や成績の向上にも寄与することが報告されている [12]。こうした技能の中でも、相手の主張の意図を正確に汲み取り的確に議論を噛み合わせる能力は、ビジネスや日常のコミュニケーションにおいても直接的に活用できる特に重要なスキルである。

この競技では発言の時間が厳格に制限されるため、複数の論点の説明や反論が同時並行で行われる。競技ディベートでは複数の話者、話者の役割、異なるステージにわたる戦略的相互作用が含まれる [2,13]。ディベーターはしばしば以前の議論に対処して戦略を適応させ、議論の進行を通じて重要な争点で優位性を確立する [14]。本競技では通常 20 個以上の主張が提示されるが、相手の議論に応答できないこと、本質的な争点から逸脱して重箱の隅をつつくような指摘に時間を費やしてしまうことは一般的に敗因とされる [15,16]。こうした競技ディベート特有の性質により、本競技における議論は複雑に交錯する傾向にあり、グループディスカッションなどの他の議論形式と比べ、より対立的かつ入り組んだ展開が見られる [17]。

そのため、議論全体を俯瞰する観点は重要な評価指標であるにも関わらず [18]、ディベータは個々の論点の詳細に集中するあまり、大局を見失うことが多い。しかも、こうした対話の破綻に気づくことは容易ではなく、俯瞰的視点が不足しているという問題意識を持つこと自体が難しいという構造的課題がある。

こうした課題を解決し、競技ディベートが有する教育的効果を最大化するためには、現在の評価手法や学習環境において複数の障壁が存在する。

第一に、既存の定量的なディベート評価手法の多くが勝敗の予測に偏重しているという点である。Ruiz-Dolz ら [19] や Hsiao ら [20] は、反論や議論の対立の構造に着目した分析を試みているものの、その評価基準は勝者予測の精度にある。しかし、実際の競技において勝利は議論の噛み合い度合いの高さと関連性が低い場合がよく見られる。相手チームの自滅や極端な主張による論破などの非建設的な議論であっても、相手側より相対的に説得力があれば勝利しえるためである。そのため、試合の勝敗だけでなく、参加者が互いの議論をどの程度効果的に噛み合わせ、建設的に対話を深められたかという点に着目した指標が必要である。

第二に、学習者が議論を俯瞰するスキルを向上させるための定量的評価や学習の方法が十分に整備されていないという点である。先述したように、競技ディベートでは毎回異なる多種多様な論題が扱われるが、現状のディベート学習では他者の試合を参考にする場合も、自身の試合を振り返る場合も、特定の論題においてどう反論すべきだったかといった、その試合での話題に深く依存した局所的な側面に意識が向きやすい。そのため、議論の構造的な組み立て方や反論の役割分担といった、論題が変わっても汎用的に活用できる知見を得にくく、次以降の試合に向けた効果的な改善が行われにくい。

1.2 反論構造の重要性

前述の評価と学習における課題は、いずれも議論の流れを俯瞰して捉える視点の欠如に起因している。評価においては、勝敗という結果や個々の発言の説得力に注目するあまり、試合全体を通じて議論がどのように展開し噛み合ったかという構造的な側面が見落とされやすい。学習においては、特定の論題で何を言うべきだったかという内容面に關心が集中するため、反論がどのタイミングでどの主張に向けられたかといった、論題を超えて応用可能な議論の組み立て方への意識が希薄になりやすい。つまり、いずれの場合も、個別の発話内容を超えて試合全体の議論構造を俯瞰的に把握する手段が確立されていないことが根本的な問題である。そこで本研究では、議論における個々の発話内容

そのものではなく、それらが互いにどう相互作用し、対話を深めているかを示す反論構造に着目する。

議論の構造的理解の重要性については、教育学や認知科学の分野で多くの研究がなされている。Jordanouら [21] は、議論スキルの発達において構造的な議論への着目が、戦略的スキルとメタ戦略的スキルの両方を向上させることを実証している。また、メタ認知研究においては、推論・議論・討論を促進することがメタ認知と自己調整学習を支援することが示されており [22]、議論構造の理解は単なる知識習得を超えた転移可能なスキルの獲得につながると考えられている [23]。これらの知見は、反論構造という議論の骨格を可視化し理解することが、論題を超えて活用できる汎用的な議論構築力の習得に寄与する可能性を示唆している。

こうした背景を踏まえると、本研究において反論構造に着目することには、大きく2つの意義があるといえる。1つ目は、議論の噛み合い度合いの定量化である。反論がどのように展開しているかという構造的特徴を定量的に解析することで、議論がどのように深められているかを客観的に評価することが可能となる。2つ目は、学習における構造的理解の促進である。個別の議論の内容から離れ、反論関係をグラフとして可視化することで、学習者は特定の論題に依存しない議論の流れを俯瞰的に把握できるようになる。これにより、他者や自身の試合の特徴を即座に把握することが可能となる。このように、構造へ着目することは、客観的で幅広いトピックに応用可能な俯瞰的議論スキルの習得に寄与すると考えられる。

1.3 本研究の目的

本研究の目的は、競技ディベートにおける反論構造を定量化および可視化することにより、試合の特徴を多角的に理解・評価するための枠組みを構築することである。本研究では、まず反論構造のどのような特徴が議論の噛み合い度合いを反映するかを明らかにし、その知見を反映した可視化手法を設計・評価するという段階的アプローチを採用し、以下の3つの取り組みを行う。

第一に、反論構造の基礎的な特性を調査するため、反論構造の定量的な特徴について分析する。具体的には、試合全体の反論構造に基づいた新たな評価指標を定義し、勝敗予測のみに依存しない議論の噛み合い度合いの定量化を試みる。ここで明らかにする反論の構造的特性は、後続の可視化手法の設計指針とする。

第二に、反論構造の分析で知見を反映し、反論のラリーや集中度合い、順序といった

構造的特徴を視覚的に把握できる反論グラフの可視化デザインを設計する。そのうえで、構造的な特徴に基づいた試合動画の探索を支援するシステム「DebaTube」を提案し、他者の試合から学ぶ際に可視化がどのように活用されるかを検証する。

第三に、DebaTubeで確立された可視化デザインを自身の試合の振り返りに応用するシステム「DebaTube Live」を構築する。参加者が自身の行った試合を試合終了後に即座に可視化することで、試合後に自身の議論を俯瞰的に振り返るプロセスにおいて可視化がどのような効果をもたらすかを検証する。

これらを通じて、評価と学習の両面から競技ディベートの教育的価値を高め、より建設的で優れた議論を展開するスキルの習得を支援することを目指す。

第2章 関連研究

2.1 計算議論学と論述文における議論構造の解析

計算議論学 (Computational Argumentation) は、人間の議論的推論プロセスを計算機的な観点から模倣・支援することを目的とした人工知能の一分野である [24,25]. この分野において、人間の議論プロセスは伝統的に識別、分析、評価、生成といった段階として捉えられてきた [26,27]. このうち、自然言語テキストから議論の構成要素を自動的に識別し、その構造を分析するタスクは議論マイニングと呼ばれ [28,29], 自然言語処理分野において活発に研究されている [30–32].

議論マイニングは一般に、議論的な談話の分割、議論単位の検出、議論関係の識別という3つのサブタスクから構成される [32]. 議論構造を推定するための第一歩は、テキストを解析可能な最小単位へと分割することである. 議論構造の理論的基盤としては、主張や前提、論拠などの議論セグメントからなる Toulmin のモデル [33] や、テキストの修辞構造を階層的に表現する Mann らの修辞構造理論 [34] が古くから知られている. しかし、自然言語処理においてこのモデルのような複雑な構造を厳密に適用することは計算コストやアノテーションの難易度の観点から困難である. 初期の研究では、入力された自然言語を「文」単位で分割するアプローチが主流であったが [35], 文脈によっては一文が必ずしも1つの議論単位と対応しないことや、文の境界が議論の構成要素の境界と一致しないことが課題とされた. こうした背景を受け、Peldszus と Stede [29] は、議論構造を計算機的に扱えるようにするため、基本談話単位と呼ばれる文法ベースの議論セグメントの概念をさらに発展させ、議論解析における最小単位として Argumentative Discourse Unit (以下 ADU) を定義した. これ以降の研究では、テキストを ADU ごとに分割することが議論マイニングの第一段階とされている [31].

こうした ADU の定義に基づき、Stab ら [36,37] は論述文を題材とした代表的な議論マイニング手法を提案した. 彼らはコーパスを主張や前提などの ADU に分割し、単位間の支持や攻撃の関係を機械学習を用いて推定した. その結果、CRF を用いた手法により議論単位の分割で F 値 0.849 を、SVM を用いた手法により各議論単位の役割分類で F 値

0.865 を達成した。また、議論単位間の支持・攻撃関係の推定については、SVM で F 値 0.733 を達成したことを報告している。

Stab らのアプローチは、議論の分割や分類、関係抽出を個別のタスクとして順次処理するパイプライン型のモデルであった。しかし、この手法には前段のタスクの誤りが後段に伝播するという課題がある。これに対し、近年では深層学習技術の発展に伴い、これらのタスクを単一のモデルで同時に学習・推論する End-to-End な議論マイニング手法が主流となりつつある。Eger ら [38] は、議論単位の抽出、役割分類、および関係推定をシーケンスラベリング問題として定式化し、LSTM などのニューラルネットワークを用いることで、人手による特徴量エンジニアリングに依存しないモデルを構築した。さらに現在では、Transformer ベースのモデル (BERT 等) を用いることで、より長期的な文脈や、ドメインを跨いだ汎用的な表現獲得が可能となっている [39]。

一方で、栗林ら [40] は大局的な論述構造の解析を目的として、接続表現の特性などを考慮した議論単位のベクトル表現を計算するモデルを提案した。実験の結果、多粒度の文脈の反映が複雑な構造を持つ論述文における論述関係認識等の精度を向上させる可能性を明らかにした。この研究は議論における文脈を低い粒度で幅広く捉えることを基本方針としているという点で、本研究と類似している。しかし Opitz ら [41] が指摘しているように、この研究では接続表現などの文法的な要素に対象が限られており、対話における反論構造を扱う本研究とは着眼点が異なる。

2.2 複数人間の対話的議論における議論の質の評価

前節では個人のモノローグ的な文章が対象とされていた研究を扱ったが、2名以上の立場が異なる者の中で交わされる対話的な議論に関して、反論構造や議論の特徴量に基づく定量的なディベートの質の評価が行われている。Zhang ら [42] は、即興型のオックスフォード式ディベートを対象とし、その成績を予測するために、議論の流れに焦点を当てた先駆的な研究を行った。彼らは、発言が論題からどれだけ逸脱したか、議論を継続させたかといったトピック関連の特徴、拍手や笑いの回数などの聴衆の反応、および発言の長さや質疑応答の戦略性といった会話特徴に根差したマクロな構造的な特徴を定量化し、議論のダイナミックな対話的側面が聴衆の意見を変えさせる力と関連することを示した。特に、勝者は敗者よりも積極的に対戦相手の論点に反応しており、こうした対話的特徴が説得力の向上に重要であることを指摘した。この研究は、より構造化されたディベートにおいて、単純な内容分析だけでなく議論の噛み合い度合いの分析が重要であること

を指摘する初期の例として位置づけられる。

Zhang らが議論の対話的構造を会話的な特徴も含めて捉えたのに対し、その後の研究では、より反論の構造に絞った分析が試みられている。Hsiao ら [20] は、オンライン上の複数ターンに渡るディベートにおける議論の質を評価する機械学習モデルを提案した。具体的には、説得を目的とした議論 [43] のコーパスで訓練し、オンラインディベート [44] のコーパスでテストする実験を行った。結果として、ターンを跨いだ反論関係を学習することで、戦略的な反論の順序付けや主張とその正当化との時間的な近接性などといった反論の構造的特徴を定量的に捉えることにより、勝者予測の精度が向上すること、相手の重要な主張への反論には、多くの理由づけが行われている主張を用いた方が効果的である可能性を明らかにした。しかし、この研究では分析対象がオンラインの掲示板のコーパスであったため、短文によるやり取りが中心であった。そのため、長文の議論の特徴は調査されていない。また、反論関係の推定では2ターン以上前の発言が考慮されておらず、反論を分析する範囲が限定されていた。

一方で、Ruiz-Dolz ら [19, 45] は、長期間特定のトピックに関する調査を行う、準備型と呼ばれる形式のディベート大会の試合の録音から作成したコーパスを対象に議論の質を分析した。結果として、Argumentation Framework [46] と呼ばれるグラフベースの議論構造のモデルを用いて、攻撃と防御的な反論の関係に基づき算出した許容可能な議論のデータを反映した手法のほうが、TF-IDF や BERT などの自然言語処理技術のみを用いた従来の手法よりも高い精度で勝者予測を行える可能性を示した。しかし、この研究では前述の通り準備型のディベートの試合が対象となっている。この形式では事前に考えた反論を型通りに提示していく、いわば作業的に最適化された議論がよく行われるため、即興型のディベートと比べ反論構造の重要性が低い。本研究は即興型のコーパスを通じてよりダイナミックな議論の流れを分析する。

これらのことから、反論構造から優れた議論パターンの特徴を定量的に捉えることによって、議論の質の評価の精度が向上していると言える。しかし、Zhang らの研究が示唆した対話的な流れの重要性は、Hsiao らや Ruiz-Dolz らの研究によって反論の構造という観点からさらに深掘りされたものの、依然としてこれまでの研究のほとんどは勝者予測に焦点を当てており、しかも限られた範囲内での相互作用しか考慮していない。そこで本研究では、試合全体の全ての反論関係を考慮に入れて大局的な観点から議論を評価することで、さらなる構造分析の可能性を探る。

2.3 反論の自動判定

特定の主張に対して、どういった反論が最も適切であるかを計算的に特定しようとする試みは、数多くなされている。Mirkin ら [47,48] は、ディベートの熟練者による 50 論題・200 スピーチを収録し、機械が議論の構成要素を 64.6%の精度で特定できることを明らかにした。これらの知見に基づき、Lavee ら [49] は、適切な議論をリアルタイムで特定する技術について研究を行った。具体的には、コーパスを 200 論題・400 スピーチへと拡張してアノテーションを付与し、その上でスピーチ中の主張を自動検出するためのいくつかのベースラインを提示した。加えて、Orbach ら [50] は、長い議論文に対する反論を生成するために、スピーチにおける反論の 55 の典型的パターンを特定した。その結果、この手法が幅広い論題に適用可能であることを示した。さらに、Orbach ら [51] は、単にスタンスが異なるだけでなく、特定のスピーチの議論に直接反論するスピーチを検出するタスクを提案した。彼らは 3,685 件のディベートスピーチからなる大規模データセットを構築し、反論関係のアノテーションを付与した。複数のアルゴリズムを検証した結果、一定の成功を収めたものの、いずれも熟練者の判断には及ばないことが示され、改善の余地があることが明らかになった。

こうした反論パターンの特定において、議論学の理論的基盤として重要な役割を果たすのが論証スキームである。Walton ら [52] によって体系化された論証スキームは、専門家の意見に基づく論証や結果に基づく論証など、人間が用いる典型的な推論パターンを分類したものである。各スキームには、その推論が成立するかを問うためのクリティカルクエスチョンが付随しており、これを用いることで、単なる論理的な整合性だけでなく、論証の質や反論の妥当性をより詳細に評価することが可能となる [53]。計算機による反論生成や判定においても、こうしたスキームの情報を利用することで、より説得力のある反論の特定が試みられている。

さらに、Wachsmuth ら [54] および Shi ら [55] は、特定の主張に対する最良の反論を特定するという、より難易度の高いタスクに取り組んだ。Wachsmuth らは、人間のアノテータによる事前評価に基づいて決定されたオンラインディベートフォーラムのコーパスから、最良の反論を検索する手法を提案した。その結果、論題にかかわらず、平均 7.6 個の候補の中から 60%の精度で最良の反論を評価可能であることが示された。Shi らは、同様のタスクに対して BERT ベースの Bipolar-encoder モデルを提案した。その結果、相手の発言から反論関係を 74.5%の精度で推定可能であることが明らかになった。

しかし、いずれの研究も反論の推定やディベートの勝敗予測に焦点が当てられており、

文脈が反映される範囲が限定的である。本研究は、試合全体の反論関係を評価することで、より広い視野で議論を分析できるようにすることを目指すものである。

2.4 テキスト提示による議論構造の理解支援

議論支援に関する研究は盛んに行われている。これらの研究の手法は、テキストを提示するものと可視化を行うものに大別される。本節では前者について概観する。Wambsganssら [56] は、学生が論理的に論述文を書けるように、議論の説得力を理解するための支援をチャット形式で動的に行うシステムを提案した。結果として、従来の筆記ツールと比較して、システムを利用した方が、ユーザが説得力のある文章を考えられることが明らかになった。しかし Guerraoui ら [57] が指摘しているように、この研究で用いられたモデルには、主張を改善する方法を説明する能力が不足している。そこで Naito ら [58] は、論述文における主張の改善案をテキストベースで提示する診断コメントの自動生成を目的としたコーパスを構築した。具体的には、即興型ディベートの試合における肯定側1番手と否定側1番手のスピーチを対象に、試合内でみられた主張の問題点を類型化した。また、Mim ら [59] は、Naito らの研究で作成されたコーパスを用いて反論の傾向を分析し、詳細な論理構造のアノテーションを行い、コーパスに追加した。今後これらの研究をもとに、主張の改善手法をテキストベースで詳細に提示するシステムが実現すると考えられる。しかし、これらの研究でのアプローチにはトピックごとの知識を反映した詳細なアノテーションが必要であり、幅広いトピックを扱う即興型ディベートに応用することは現実的でない。本研究は、トピックの知識に依存しない一般的なアプローチによる議論評価を行うものである。また、これらの研究では反論の質を判断する際に他の反論の内容は一切考慮されず、反論文と被反論文の内容のみが反映されている。本研究は、より大局的な観点から反論を評価するものである。

2.5 可視化による議論構造の理解支援

可視化による議論支援の研究では、対象とするユーザの専門性や分析の粒度に応じて、幅広いアプローチが提案されている。El-Assady ら [60] は、複数人の会話における話者の挙動やトピックの推移を、沈殿物の層のような視覚的メタファを用いて表現する ConToVi を提案し、誰がどの話題で主導権を握ったかの把握を支援した。South ら [61] はこれを政治討論会を対象に一般ユーザ向けに展開し、討論の探索や分析を支援するツール DebateVis

を提案した。評価実験の結果、システムの使用がどの政治家がどの発言を行ったかへの興味を促すことが明らかになっている。また、Ullmannら [62] は、オンライン討論における集合知に着目し、議論の健全さや参加者の発言のバランスを可視化するダッシュボードを実装している。

また、専門家のユーザを対象に議論の論理構造や詳細な分析を支援する研究が行われている。Carneiroら [63] は、テレビ討論番組の映像を対象に、複数ユーザによる同期的な協調分析を可能にするツール「Deb8」を提案した。このツールは、映像中の発言の引用、意見、質問、外部エビデンスを論証グラフとして結びつけ、議論構造を可視化するインタフェースを提供する。また、各要素の関連性や信頼性を協調的にタグ付けする機能を備えている。3人1組のグループ3組によるユーザスタディを通じて、ツールの設計を評価し、協調的な議論分析における課題を明らかにした。Sevastjanovaら [64] は、専門家が議論データを分析し、重要な洞察を抽出してレポートを作成するプロセスを支援する VisInReport を提案した。5名の政治学や言語学の学者によるユーザスタディの結果、システムが分析とレポート作成の反復プロセスを効率化し、深い洞察の獲得を促進することを明らかにした。また、Gruberら [65] は、Inference Anchoring Theory (IAT) に基づき、議論の動的な構造を可視化するツール「DebArgVis」を提案した。QT30 コーパス (BBC の討論番組データ) を用いたケーススタディを通じて、発言間の論理的結合やトピックの推移を追跡する上でツールが有用であることを示した。こうした流れを受け、Moonら [66] は、テレビ討論番組を対象に、議論セグメントの識別と意見の比較を支援する可視化システムを提案した。このシステムでは建築物や化合物といった視覚的メタファを採用することで、非構造的な議論データを構造的に提示している。実際の討論データを用いたシナリオベースのユーザ実験の結果、共起行列を用いた可視化が議論の全体構造や頻度の把握に、サークルパッキングを用いた可視化がトピックごとの参加者の意見や立場の比較にそれぞれ有効であることが明らかになった。

さらに、競技ディベート特有の戦略的な反論に着目した研究として、Chenら [67] による Conch が挙げられる。彼らは、議論の「争点」と「反論戦略」を階層的に抽出し、スパイラル状のタイムラインと積み上げ棒グラフを用いて議論の展開と戦略の共起関係を可視化した。特筆すべきは、動画視聴やテキスト閲覧との比較実験において認知負荷理論に基づく評価を行っている点である。その結果、可視化システムの利用が分析に伴う外在的負荷を低減させるだけでなく、学習に寄与する生成的負荷を有意に高めることを示し、ツールが深い洞察の獲得を促進することを定量的に実証した。

また、教育分野では、議論構造の可視化による論述文のライティングの支援が行われている。Wambsganss ら [68] が機械学習により推定した議論単位間の支持関係をグラフ化することで、学生が論述文の論理構造を把握する支援を行った。また、システムの利用実験により学生の文章の説得力が向上する可能性を明らかにした。Xia ら [69] は、他者の論証構造のグラフ提示により、ユーザの意見の説得力が向上する可能性を明らかにしている。

しかし、いずれの研究も議論構造の可視化手法の提案や、分析ツールとしての機能的な有用性の検証が主であり、Conch [67] のように学習効果を定量的に示した例はあるものの、反論構造グラフの特徴がユーザにどう解釈されるかといった可視化を活用するプロセスに関する分析は十分に行われていない。また、ディベートを行った者に自身の試合の可視化を提示した研究は管見する限り存在しない。本研究は、ディベート学習のための他者の試合の動画探索と自分が参加した試合の振り返りという実践的なタスクを通じて、ディベーターが可視化から学びを得るプロセスに着目するものである。

第3章 議論の噛み合い度合いの評価のための反論構造の定量化

1章で述べたように、競技ディベートは教育的に意義深い取り組みであり、その効果を最大化するためには全参加者がどの程度議論を深められたかという議論の噛み合い度合いに着目する必要がある。しかし、既存の定量的評価手法の多くは勝敗予測に偏重しており、実際には相手側の自滅や非建設的な論破によって勝利する場合もあるため、勝敗のみでは議論の噛み合い度合いを十分に捉えられない。また、現状でもジャッジからの講評や教育者からの指導を通じて、建設的な対話の重要性や議論の深め方について指摘を受けることはあるが、こうしたフィードバックはほとんどが定性的なものである。こうした評価は個人の知識や経験に依存するため、評価者が精通していない分野では的確な指摘が難しい場合もあり、論題によらず評価の質を担保することは難しい。したがって、論題に依存せず、かつ教育的に意義のある勝敗から独立した評価指標が求められる。

ここで、再反論や再々反論といった反論のラリーの深さや、2人以上前の発言への反論の割合といった反論の構造的な特徴は、両チームの議論がどの程度噛み合い、どのように展開されたかという議論の噛み合い度合いを反映すると考えられる。また、こうした構造的な特徴は議論の内容そのものではなく主張間の関係性に基づくため論題に依存せず、評価に活用できれば幅広い試合に対応できると考えられる。

そこで、本章では競技ディベートにおける議論の噛み合い度合いを反論構造から定量的に評価する手法を構築する。具体的には、「競技ディベートにおいて反論の構造的な特徴は議論の噛み合い度合いを示すか」という問いを検証するため、試合全体の反論構造の特徴量を定義し、専門家とLLMのハイブリッド手法によって得られる議論の噛み合い度合いとの関連性を評価する。

3.1 反論構造のモデル化

3.1.1 用語の定義

反論構造を分析するために、まず重要な用語を定義する。2.1 節で述べたように、議論構造の解析においては Peldszus ら [29] により定義された ADU が分析の最小単位として広く用いられている。しかし、即興的な発話が行われる競技ディベートの解析においては、Stab ら [36] のような多層的な分類 (Major Claim, Claim, Premise 等) を適用することは困難である。実際、Slonim ら [70] は、計算的な議論構造の解析において複雑なカテゴリは不適當であると指摘している。そこで、本研究における ADU は、「理由と主張のペア」、または「独立した主張」のいずれかを指すというシンプルな定義を採用する。この柔軟な定義は、主張が即座に裏付けを伴うとは限らないという競技の特性にも対応している。この ADU は本研究における基本的な分析単位である。

ここで、この競技には相手のスピーチ中に、POI(Point of Information) とよばれる質問を短い時間で行うことができるというルールがある。通常、これは時間的制約により単一の主張となるため、個別の ADU として扱う。これは、先述の単体の主張を 1 つの単位として扱う ADU の定義と整合する。

反論は、対戦チームによってなされた特定の議論に応答する発言として定義する。相手の主張を予測することや、特定の発言に言及することなく対立する立場を提示することは、反論とはみなされない。本モデルでは、反論は ADU 間の有向エッジとして表現される。また、ノードは時系列に並べられ、肯定側の ADU が左側に、否定側の ADU が右側に配置される。

3.1.2 反論の構造的特徴に基づく仮説

本章の問いである「競技ディベートにおいて反論の構造的特徴は議論の噛み合い度合いを示すか」を調査するために、説得的議論と反論構造に関する先行研究に基づいた 4 つの仮説を立てる。対話の構造を直接扱った研究はほとんどないが、議論の説得力に関する研究から、参考になる知見が得られている。Hsiao ら [20] は、議論とその正当性の説明の間の時間的な間隔が短いほど説得力が高まることを明らかにした。このことから、反論のタイミングと位置関係が議論の噛み合い度合いに影響すると考えられる。また、Hsiao らは反論の順序を戦略的に考えることが説得力に寄与することを示した。複数のトピックが同時に扱われる競技ディベートでは、各トピック内で相手の議論の順序を維持するこ

とが論理的な流れを保つのに役立つため、反論の順序を揃えることは優れた対話において重要であると考えられる。さらに、Ruiz-Dolz ら [19] は Dung の抽象議論フレームワークを応用し、議論が攻撃された場合には反撃が必要であるという原則に基づいて議論の質を評価することに成功した。ただし、Dung のモデルは本来、単一の論者による議論構造を想定している。競技ディベートのようなチーム間の対話においては、単発の再反論だけではなく、再々反論といった深い反論のラリーが建設的な議論を生むと考えられる。

以上のことから、反論の遠さ、位置関係、順序、ラリーという構造的特徴に着目し、以下の仮説を立てた。

H1(Distance): 2人以上前のスピーチへの反論の集中は、直前のスピーチの発言が無視され、議論の流れが絶たれていることを示唆する。そのため、このような遠い反論の頻度が高いほど議論の噛み合い度合いが低い。

H2(Interval): 同じ発言への反論の間隔が大きく空いている場合、やり取りの一貫性が弱まる。そのため、同じ発言への反論の間隔が大きいほど議論の噛み合い度合いが低い。

H3(Order): 相手の発言のと順序が揃っている反論は一貫性が高く、議論の噛み合い度合いが高い。

H4(Rally): 再反論が連続し、反論のラリーが長く続く場合、議論が深められており議論の噛み合い度合いが高い。

3.1.3 反論構造の特徴量の定義

3.1.2 項の仮説 H1–H4 に対応する特徴量を定義する。まず、特徴量の計算に用いる要素と関数を以下に示す。なお、これらの計算は試合ごとに独立して行われる。

- ADU_i : i 番目の ADU
- $ADUs$: 全 ADU のリスト
- $rebs$: 全反論の集合
- $speeches$: 全スピーチのリスト
- $reb[i, j]$: ADU_i から ADU_j への反論
- $len(array)$: 配列 $array$ のサイズ
- $speechOf(i)$: ADU_i を含むスピーチのインデックス (1 始まり)

- $adusIn(i)$: i 番目のスピーチ内の ADU の集合

次に、仮説 H1–H4 のそれぞれに対応する 4 つの反論構造の特徴量 ($Distance$, $Interval$, $Order$, $Rally$) を定義する。これらはアルゴリズム 1 から 4 により計算される。

Algorithm 1 Distance : 遠い反論の割合 (H1 に対応)

```

1:  $count \leftarrow 0$ 
2: for all  $\{reb[src, dst] \in rebs \mid speechOf(src) \geq 4\}$  do
3:    $dist \leftarrow speechOf(src) - speechOf(dst)$ 
4:   if  $(speechOf(src) = len(speeches) \text{ and } dist \geq 2)$  or  $dist \geq 3$  then
5:      $count \leftarrow count + 1$ 
6: return  $\frac{count}{len(rebs)}$ 

```

Algorithm 2 Interval : 同じ発言への反論の間隔 (H2 に対応)

```

1:  $interval \leftarrow 0$ 
2: for  $i \leftarrow 1$  to  $len(speeches)$  do
3:    $dsts \leftarrow \{dst \mid reb[src, dst] \in rebs, speechOf(src) = i\}$ 
4:   for all  $d \in dsts$  do
5:      $rebsToSameADU \leftarrow \{reb[src, dst] \in rebs, dst = d, speechOf(src) = i\}$ 
6:      $srcs \leftarrow \{src \mid reb[src, dst] \in rebsToSameADU\}$ 
7:     if  $len(srcs) > 2$  then
8:        $interval \leftarrow interval + \frac{\max(srcs) - \min(srcs) - 1}{len(adusIn(i)) - 2}$ 
9: return  $interval$ 

```

Algorithm 3 Order : 反論の順序の一貫性 (H3 に対応)

```

1:  $crosses \leftarrow 0$ 
2: for all  $reb[src_i, dst_i] \in rebs$  do
3:   for all  $reb[src_j, dst_j] \in rebs$  where  $j \neq i, speechOf(src_i) = speechOf(src_j)$  do
4:     if  $(src_j - src_i)(dst_j - dst_i) < 0$  or  $src_i = src_j$  then
5:        $crosses \leftarrow crosses + 1$ 
6:       break
7: return  $\frac{len(rebs)}{crosses}$  if  $crosses > 0$  else  $-1$ 

```

Algorithm 4 Rally : 反論のラリー度合い (H4 に対応)

```

1:  $chains \leftarrow 0$ 
2: for all  $reb[src_i, dst_i] \in rebs$  do
3:   for all  $reb[src_j, dst_j] \in rebs$  where  $j \neq i$  do
4:     if  $dst_i = src_j$  then
5:        $chains \leftarrow chains + 1$ 
6: return  $\frac{chains}{len(speeches) \cdot len(rebs)}$ 

```

3.1.4 反論構造の特徴量の解釈

本項では、各構造的特徴がどのように計算され、解釈されるかを述べる。図 1 は反論構造の可視化事例である。各図において、黒い枠のブロックはスピーチ内で時系列に配置された ADU を表す。赤い線は反論リンクを示し、下端が反論の発信元を指し、上端が標的を指している。

図 1a: 反論の遠さ (*Distance*)。これは遠い反論の割合が大きく、議論の噛み合い度合いが低い事例である。ここでは肯定側 3 番手から反論が 4 回行われているが、3 つは 2 人前のスピーチに反論し、1 つは直前のスピーチに反論している。3.1.3 項での定義によれば、2 人以上前への反論は「遠い反論」とみなされる。したがって、このグラフでは 4 つの反論のうち 3 つが遠い反論と判断され、*Distance* スコアは 0.75 となる。*Distance* が高いほど過去の議論への反論が多く、直前のスピーチが無視されていることを示すため、仮説 H1 に基づき議論の噛み合い度合いが低いと推定される。

図 1b: 同じ発言への反論の間隔 (*Interval*)。これは同じ発言への反論の間隔が大きく、議論の噛み合い度合いが低い事例である。ここでは同じスピーチから ADU_1 に対し 3 回反論が行われている。これらの反論は最初のものが ADU_2 から、最後のものが ADU_6 から提示されており、それらの間には 3 つの ADU がある。そのため、これらの反論群の間隔は 3 として計算される。*Interval* が高いほど同じ発言への反論が散在しており、議論が十分に整理されてないことを示すため、仮説 H2 に基づき議論の噛み合い度合いが低いと推定される。

図 1c: 反論の順序の対応度 (*Order*)。これは反論の順序の対応度合いが低く、議論の噛み合い度合いが低い事例である。ここでは、同じスピーチからの 2 つの反論のエッジが交差している。すなわち、 ADU_5 および ADU_6 が、これらより先に提示された ADU_3 のターゲットよりも前の ADU に反論している。これは反論の順序が一致していないことを意味する。*Order* スコアは反論の数を交差の数で割ったものとして定義される。*Order* が低いほど反論の交差が多く組織的な反論が十分なされていないことを示すため、仮説 H3 に基づき議論の噛み合い度合いが低いと推定される。なお、交差が存在しない場合は -1 を返す。また、同じ発言から複数の反論が行われる場合は話題が交錯していると考えられるため交差として扱った。

図 1d: 反論のラリー度合い (*Rally*)。これは反論のラリーが続いており、議論の噛み合い度合いが高い事例である。 reb_i は i 番目の反論のエッジを表している。ここでは reb_1 の反論先が reb_2 の反論元となり、さらに reb_2 の反論先が reb_3 の反論元となっており、

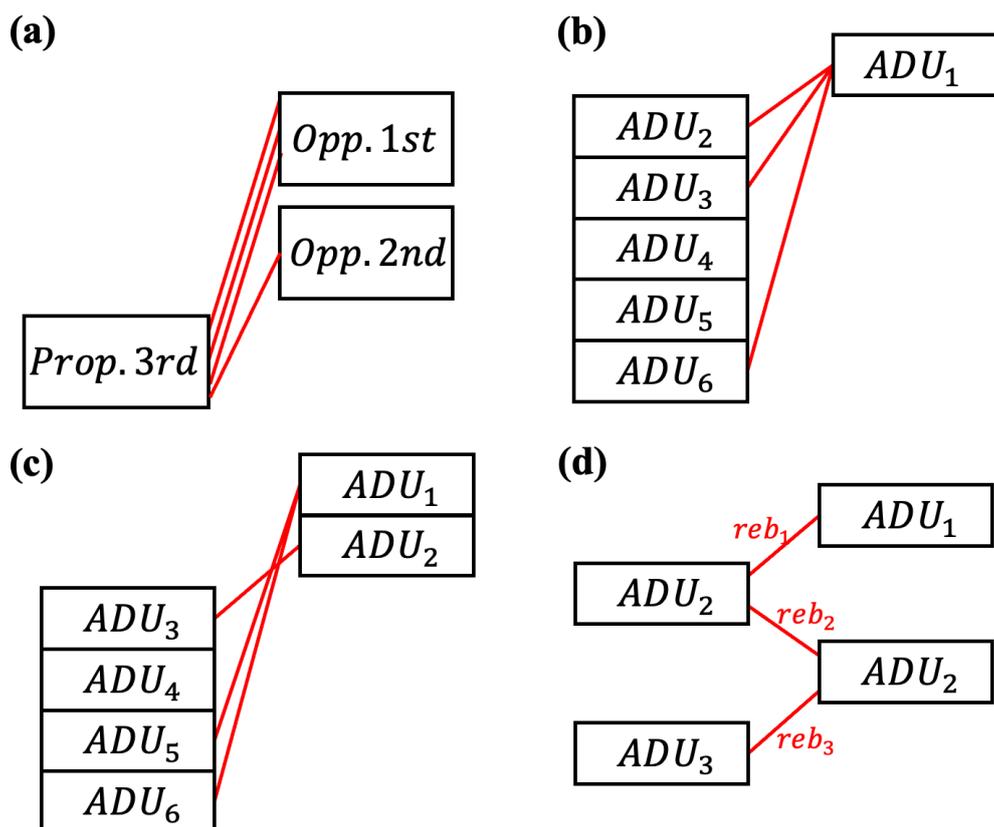


図 1: 4つの構造的特徴が顕著な事例の可視化. (a) *Distance*: 2つ以上前のスピーチへ反論が集中している (この例のみ ADU はスピーチごとに結合して表現されている) (b) *Interval*: 同じ発言への反論の間隔が大きく空いている (c) *Order*: 反論の交差が頻繁に発生している (d) *Rally*: 再反論の連鎖が発生している

反論の連鎖を形成している. アルゴリズム 4 では, このように, ある反論のターゲットが他の反論の反論元と一致しているような反論のペアの数を *chains* としてカウントしている. この事例では $chains = 2$ となる. これは, 反論グラフにおける全てのパスについて, 各パスの長さから 1 を引いた値の総和に相当する. このような *Rally* のスコアが高い試合では, 仮説 H4 に基づき議論が積極的に深められており議論の噛み合い度合いが高いと推定される.

3.2 コーパスの構築

3.2.1 データ構造

仮説を検証するため, 一定以上のディベートの経験がある学生による, 練習試合および大会予選における 2 チーム制即興型競技ディベートの動画を 20 本収集した. すべての試合で日本の高校生または大学生が英語でディベートを行っている. このような英語を母

語としない参加者の、中級レベルの試合を中心に選定した理由は、優れた議論だけでなく、弱い議論や欠陥が含まれる討論も調査対象とするためである。これにより、幅広い特徴を持つ議論の特性の分析が可能となった。

収集した動画は音声認識モデル (whisper-large-v2) を用いて書き起こされた。また、書き起こしテキストは、3.1.1 項の定義に基づき、ADU への分割と反論関係のアノテーションが行われた。2.1 節で述べたように、議論構造の推定にはいくつかの計算的手法が存在する。しかし、競技ディベートでは主張が複数の非連続的な文にまたがって表現されたり暗示的にしか述べられないことが多く [71]、こうした暗示的な前提の反映は依然として困難だとされている [72]。そのため、3 年以上の競技ディベートの経験を持つ著者により手作業で行った。

反論関係は、同じトピックに関する相手の議論の特定の側面に対する応答として定義された。各 ADU には試合ごとの時系列順にインデックスが付与され、反論関係はソースとターゲットの ADU インデックスの組 (tuple) として記録された。また、POI (Point of Information) に対応する ADU インデックスのリストも作成した。

3.2.2 議論の噛み合い度合いの評価方法

3.1.3 項で定義した構造的特徴を評価するため、各試合における議論全体の議論の噛み合い度合いを定量化した。その際、3 名の評価者 (人間の専門家、人間の非専門家、および LLM (大規模言語モデル)) を用いて、各試合の評価を行った。

専門家の評価者は、即興型英語ディベートにおいて 10 年以上のジャッジ経験がある。非専門家の評価者は、ディベート経験が 3 年あるがジャッジ経験はない。3 人目の評価者である LLM には OpenAI の o3 モデルを使用した¹。

ディベートの評価は伝統的に人間のジャッジに依存してきたが、Jingcong ら [73] および Liu ら [74] が LLM がディベートの質をある程度妥当に評価できることを明らかにしている。同時にこれらの研究では LLM は後半のスピーチを軽視する傾向にあるといったバイアスがあることも示されており、LLM を人間による評価の完全な代替とすることは難しいが、専門家による評価には、1 試合あたり約 3 時間という膨大な時間を要するため、全 20 試合に対して実施することは現実的ではない。そこで、本章を探索的質的研究と位置付け、専門家と LLM の評価を融合したハイブリッドアプローチにより議論の噛み合い度合いを定義することとした。

¹使用したプロンプトの詳細は <https://osf.io/ceugg> で公開している

全ての評価者は、各ディベート試合の書き起こしと動画の両方を確認し、一貫した基準（人間には実験ガイドライン、LLMには同等のプロンプト）に基づき評価を行った。具体的には、「両チームに関して試合全体を通して議論が噛み合っていた」という項目について、4段階のリッカート尺度（そう思う (4)、ややそう思う (3)、ややそう思わない (2)、そう思わない (1)) を用いて回答した。LLM の評価については、同じプロンプトを用いて各試合につき 5 つの出力を生成し、その最頻値をその試合の最終的な評価値とした。90% の試合において 5 回中 3 回以上の評価が一致したため、このアプローチの信頼性は許容できると判断した。

評価者間信頼性を評価するため、先述の 4 段階評価を、否定的 (2 以下) と肯定的 (3 以上) の二値分類に変換し、コーエンのカッパ係数を算出した。二値分類を用いたのは、「そう思う」と「どちらかというと思う」といった比較的重要な不一致の影響を軽減し、サンプルサイズに対してより堅牢な信頼性推定を行うためである。分析の結果、専門家と LLM の間には中程度の合意 ($\kappa = 0.490$, 4 段階分類での完全一致率は 60%) が見られた一方で、非専門家は専門家 ($\kappa = 0.175$) と LLM ($\kappa = -0.056$) と一致率が低く、非専門家の判断が不適切であることが示唆された。したがって、以降の分析からは非専門家の評価を除外し、各試合の最終的な議論の噛み合い度合いの評価は専門家と LLM の評価の平均値として評価した。

3.3 結果

本節では、3.1.3 項で定義された反論構造の特徴量と専門家と LLM によって評価された議論の噛み合い度合いの関連性に関する結果を述べる。

3.3.1 反論構造の特徴量と議論の噛み合い度合いの相関

仮説を検証するため、3.1.3 項で定義した反論構造の特徴量を算出し、回帰分析を通じて特徴量と評価者の評価との関係を調査した。また、モデルの汎化性能を評価するため、20 サンプルのデータセットに対して leave-one-out 交差検証を採用した。特徴量の値は、各特徴量の 20 試合中の最大値で割ることで正規化した。

表 1 に示すように、重回帰分析 (Multiple Linear Regression) が最も良い性能を示し、RMSE は 0.509、MAE は 0.409 であった。予測値と評価者の評価との相関係数は 0.609 であり、中程度の正の相関が見られた。標準偏差 (SD) が 0.688 であることを考慮すると、

表 1: 回帰モデルの性能比較

Model	RMSE	MAE	Correlation Coefficient
Multiple Linear Regression	0.509	0.409	0.609
Ridge Regression	0.562	0.447	0.508
Lasso Regression	0.511	0.412	0.589

表 2: 回帰係数と特徴量の重要度

Feature	Multiple Linear	Ridge	Lasso	Importance (%)
<i>Distance</i>	1.172	0.530	0.915	28.8
<i>Interval</i>	1.030	0.611	0.914	25.3
<i>Order</i>	0.216	0.007	0.000	5.3
<i>Rally</i>	1.656	0.671	1.319	40.7

RMSE/SD が 0.740 であることは、主観評価タスクにおいて実用的に許容可能な予測性能であることを示唆している。

3.3.2 特徴量の重要度の分析

回帰係数の分析により、3つのモデル間で明確なパターンの違いが明らかになった。表 2 は回帰係数と特徴量の重要度を示す。この表から、*Rally* 特徴量が 40.7% と最も高い重要度を示し、次いで *Distance* (28.8%)、*Interval* (25.3%)、*Order* (5.3%) の順となったことが分かる。重回帰分析では、*Rally* (1.656)、*Distance* (1.172)、*Interval* (1.030) において特に強い係数が示された。特筆すべき点として、Lasso 回帰は *Order* 特徴量を完全に削除 (係数 = 0.000) しており、その特徴選択の特性を示している一方、Ridge 回帰はこの特徴量に対して非常に小さな係数 (0.007) を維持した。

3.4 分析と考察

3.4.1 評価者間信頼性

3.2.2 項で示した評価者間信頼性に基づくと、二値分類の低いカッパ係数および非専門家と他の評価者間の評価傾向の違いは、非専門家が異なる評価基準を適用していた可能性を示唆している。専門家と LLM は非専門家よりも厳しい評価傾向を示しており、LLM の評価は非専門家よりも専門家の評価パターンに近いことが示唆される。

専門家と非専門家間の評価基準の違いを明らかにするため、自由記述の評価理由 (3.2.2 項目を参照) の定性分析を行った。その結果、専門家は議論の明確性や論理的一貫性を

重視した一方、非専門家は分析の網羅性や多角的な視点を優先していたことがわかった。例えば、高齢者介護に関するディベートにおいて、専門家は曖昧なスタンスや不十分な例示を理由に低いスコアを付けたが、非専門家は多様なステークホルダーの視点を網羅しているとして高く評価した。このことは、信頼性の高いディベート評価には、単なるディベート経験だけでなく、豊富なジャッジ経験が必要であることを示している。

LLMは非専門家よりも専門家の評価と良い整合性を示しているものの、その相関は中程度 ($\tau = 0.400$) にとどまり、専門家の判断や構造的特徴に基づく評価を完全には再現できていない。LLMはディベート全体ではなく特定のスピーチを評価する傾向がある。例えば、決闘の合法化に関する試合において、専門家は肯定側第一スピーカが重要な前提を説明できず、試合を通じて反論され続けたため「どちらかというそう思わない」と評価した。一方で、LLMは序盤のスピーチにおける反論の多様性に基づいて「そう思う」と評価した。これらのことから、反論構造に基づく指標は、試合全体の反論関係を分析することでLLMが見逃してしまう対話的な構造を捉え、より堅牢な評価を行える可能性がある。

3.4.2 モデルの精度に関する詳細分析

モデルの精度について、3.3.1項で述べた結果は、H4 (*Rally* に関連) が40.7%という最も高い重要度で強く支持されることを示している。さらに、H1 (*Distance* に関連) とH2 (*Interval* に関連) は、それぞれ28.8%と25.3%の重要度で中程度に支持されている。一方で、H3 (*Order* に関連) は5.3%の重要度にとどまり支持は限定的で、Lasso回帰では完全に排除された。

これらの結果が生じた背景を分析するために、反論構造の可視化を通じて、これらの構造的特徴の詳細な考察を行う。図2は反論構造の可視化グラフである。長方形のノードがADUを、エッジが反論を表しており、赤色は肯定側の要素、青色は否定側の要素を示している。

第一に、*Rally*はH4との最も強い整合性を示し、この仮説が支持される傾向にあることを示唆している。図2(a)の試合は、最大の*Rally*値を持ち、最悪の*Interval*および*Order*スコアであったにもかかわらず、4点満点中3.0の評価を得た。これは、反論が十分に整理されていなくても、反論のラリーが長く続ければ議論が十分深められている可能性を示唆している。

第二に、*Distance*はH1を中程度に支持する傾向がみられた。図2(b)は、否定側の第一

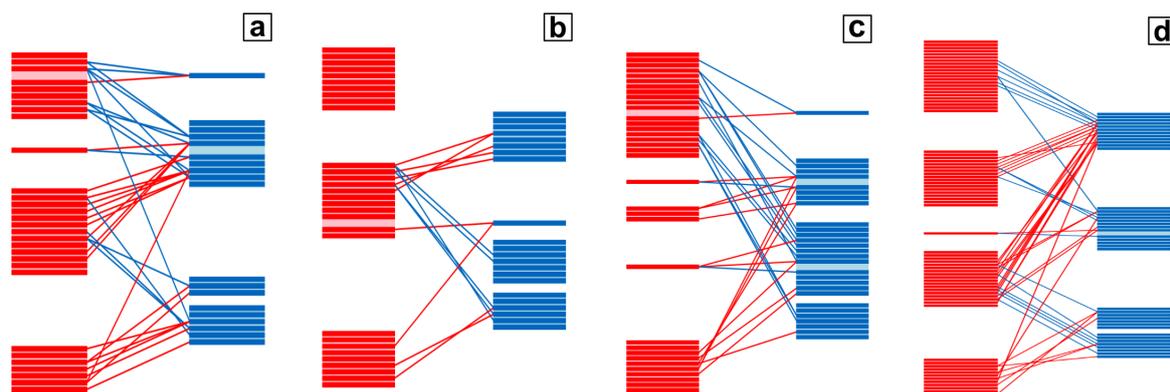


図 2: (a) 最大の *Rally*, (b) 最小の *Distance*, (c) 最大の *Distance*, (d) 3 番目に大きい *Order* の値を持つ反論構造の可視化

スピーカが全く反論を行わなかったにもかかわらず、最小の *Distance* と最高の評価 (3.5) を獲得した注目すべき事例である。これは継続的な反論が重要であることを示唆している。さらに、この特徴量は議論の欠陥も捉える。図 2(c) は最大の *Distance* 値 (2.5) を持つ試合を示しており、肯定側の第二スピーカがほとんど発言せず、否定側が第一スピーカへの反論に集中せざるを得なくなり、対話が崩壊している。これらのことから、*Distance* は対話の優れた部分と欠陥の両面を反映できると言える。

第三に、*Interval* は H2 と中程度の整合性を示した。例えば、最小の *Interval* を持つ試合は、平均的な *Rally* のスコアにもかかわらず良好な評価 (3.0) を得た。これは、チームが同じ議論に対して複数の攻撃を集中させる場合、反論の少なさを補えることを示唆している。

最後に、*Order* は H3 との整合性が最も低く、正則化手法によって削除された。交差が多い個別のスピーチが全体のスコアを不釣り合いに下げってしまうため、モデルの効果が低くなった。例えば、図 2(d) の試合は最高評価 (3.5) を得たが、*Order* スコアは 3 番目に悪かった。これは、交差の少ない強力な序盤のスピーチ（特に否定側第一スピーカと肯定側第二スピーカ）が、交差の多い後半のスピーチによって影が薄くなってしまったためである。スピーチごとに交差の重み付けを行うといったモデルの改善が必要だと考えられる。

3.5 小括

本章では、即興型競技ディベートにおける建設的な議論の特性の基礎的な調査を目的に、反論のラリーや順序、間隔などの反論構造の特徴を定量化した。具体的には、即興

型競技ディベート 20 試合分の ADU 分割と反論関係のアノテーションを行い、計算議論学の基盤となるコーパスを構築したうえで、反論の遠さ (*Distance*)、同じ発言への反論の間隔 (*Interval*)、反論の順序の対応度 (*Order*)、反論のラリー度合い (*Rally*) という 4 つの構造的特徴量を定義した。これらの特徴量と、専門家と LLM のハイブリッド評価によって得られた議論の噛み合い度合いとの関連性を検証することで、勝敗判定から独立し、かつ論題に依存しない汎用性の高い評価指標の基礎を構築した。特に、反論のラリーの深さが議論の噛み合い度合いに強く寄与するという知見は、ディベート教育において「相手の反論に対してさらに応答を重ねる」ことの重要性を実証的に裏付けるものである。さらに、専門家・非専門家・LLM の評価を比較した結果、LLM が非専門家よりも専門家に近い評価傾向を持つことが明らかになった。この発見は、従来評価者に大きな負担をかけていたディベートの質的評価のアノテーション作業を大幅に効率化できる可能性を示唆している。

第4章 DebaTubeの提案と評価

3章では、反論構造の定量的な指標を用いることで、議論の噛み合い度合いを客観的に評価できる可能性が示唆された。特に、再反論のラリーの頻度 (*Rally*)、遠い反論の割合 (*Distance*)、同じ主張への反論の間隔 (*Interval*) という3つの特徴量が議論の噛み合い度合いを評価するうえで重要な役割を果たすことが明らかになった。また、反論の順序は関連する特徴量である *Order* が議論の推定に有効であるとは示されなかったが、3.1.2項で述べた Hsiao ら [20] の分析からも、依然として議論の戦略性に関わる重要な要素であると言える。これらの知見は、ディベータが試合の質を俯瞰的に把握する際に、どのような構造的特徴が使えるかを示している。

しかし、ディベートの学習者が議論の構造や建設的な反論の方法を深く理解するためには、数値的な評価だけでなく、実際に優れた議論が展開されている試合を参照することが不可欠である。特に、味方に反論が集中した際にどうカバーするか、チーム内で反論先をどう分担するかといった構造的な側面を学ぶには、他者の試合を観察することが有効である。ここで、3章で明らかにした構造的特徴を視覚的に捉えられるようになれば、ディベートの学習に有用であると考えられる。

しかし、Debate Motions¹などの既存のディベートの試合動画を検索するためのサービスでは、論題や大会名といった試合の表層的な情報に基づく検索に限られており、反論のラリーが活発な試合や反論のバランスが取れた試合といった、試合の構造的な中身を手がかりに動画を探索する手段が存在しない。その結果、学習者は反論が入り組んだ複雑な試合、特定の特徴を持つ反論群が機能した試合などといった自身が学びたい局面を含む動画を探すことができず、反論構造を意識した学習が困難となっている。

そこで本章では、「試合での反論構造の可視化は動画探索においてどのような役割を担うか」という問いを立て、これを検証するため、3章で明らかになった構造的特徴を視覚的に把握できる可視化手法をデザインし、試合全体の反論構造を可視化することで各試合の特徴に基づく動画探索を支援するシステム「DebaTube」を提案する。本システムは、次節で述べる LLM を用いた自動推定手法により得られた反論構造グラフを、動画と紐付

¹<https://debatedata.io/>



図 3: 反論構造の推定手法の概要

けて提示することで、俯瞰的な視点からの動画探索を実現するものである。本章では、反論構造の自動推定手法、可視化手法の設計、システムの実装、およびディベート経験者を対象とした評価実験の結果について述べる。

4.1 LLM を用いた反論構造の自動推定手法

これまでの研究では、反論構造の判定は手作業で行われてきた。しかし、この手法では1試合あたりに3時間程度かかり、コーパスの大規模化やリアルタイムの推定が現実的でなく、拡張性やディベート学習での実用性、また、アノテータの主観への依存という課題があった。そこで今回、YouTube 上に動画がアップロードされている試合を対象に、LLM を用いて試合の音声データから反論構造のグラフを推定し取得する手法を構築した。図3に概要を示す。

まず YouTube 上の動画の音声データから、音声認識モデルを用いて文字起こしのテキストデータをセグメント単位（音声認識モデルが音声内の音響的な区切りに基づき決定した単位）で取得する。次に、話者識別モデルを用いて各セグメントの発話者の分類を行い、同じ話者が連続しているセグメントをまとめてスピーチとする。さらに、各スピーチに含まれる連続するセグメントをまとめ、ADU とする。2つのスピーチ間の ADU のペアを、1つの反論関係として扱う。

セグメント群 {segments} を ADU 群にグルーピングするプロンプトは下記のとおりである。

Regroup the given segments into argumentative units of 1 to 5 segments each and return the list of the first segment's id in each unit.

Argumentative units are elementary argumentation factors, such as claims, cases, and

rebuttals.

Given segments: {segments}.

その結果、各スピーチは下記のような ADU のリストとして表現される。このスピーチを、スピーチの ID である i を用いて {speech[i]} と表現する。ここで、スピーチの ID は肯定側が 1, 3, 5, 8 となり、否定側が 2, 4, 6, 7 となる。

```
[
  "1:Violence isn't just physical ...",
  "2:Violence comes from an empty ...",
  "3:Because you do not know if ...",
  ...
]
```

最後に、スピーチ間の反論関係を取得するため、あるスピーチ ID の i について、相手側の ID を j とし、その反論関係となる ADU のペアを求める。ここで、 $j < i$ とする。

Identify all rebuttals present between the source speech and the target speech.

Rebuttals are direct responses to the opponents' arguments, typically starting with a rephrasing of the argument they are addressing.

Source speech: {speech[i]}

Target speech: {speech[j]}

以上の処理により、ある ADU に対する反論である ADU をペアとし、そのエッジを反論として扱い、そのノードとエッジの群をもって反論構造とする。

ここで、ADU 間の反論関係の推定において LLM が一度判定したものを全て採用すると手作業の時と比べ著しく精度が低下するという問題が生じた。具体的には、Precision が 0.127, Recall が 0.256, F1 Score が 0.167 となっていた。そのため、反論関係のペアを同じプロンプトで 10 回生成し、閾値以上の回数リストに含まれていた ADU のペアのみを反論として採用する手法を採った。この閾値を N と表現する。

図 4 は、10 回中 N 回以上検出された ADU のペアのみを採用する条件で、3 章で作成し

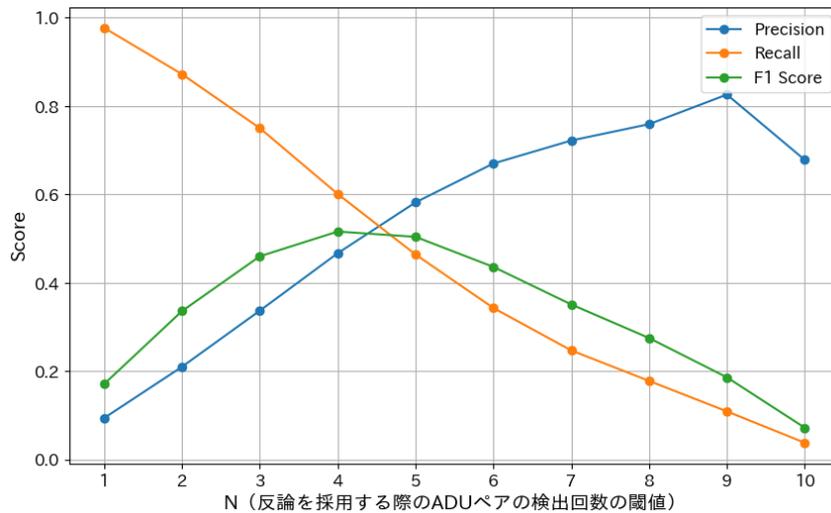


図 4: 反論を採用する際に用いた ADU ペアの検出回数の閾値 (N) と精度の関係

たコーパスを正解データとした際の反論元の推定精度である。図中の縦軸は各スコア，横軸は N（反論を採用する際に用いた ADU ペアの検出回数の閾値）である。本結果より，N=4 の時最も F1 スコアが高かったため，本システムでは N=4 の条件で推定したグラフを採用した。

4.2 反論構造の可視化のデザイン

3章で重要性が示唆された反論構造の特徴量 (*Rally*, *Distance*, *Interval*, *Order*) を視覚的に把握できるよう，反論構造の可視化グラフをデザインした。図 5 に初期デザインと最終デザインの比較を示す。本節では，初期デザインからの改善点や各特徴量が視覚的に把握できるよう可視化デザインを工夫した点を述べる。

Distance. *Distance* は，2人以上前のスピーチへの反論の割合を表す特徴量である。この特徴量を視覚的に把握しやすくするため，肯定側と否定側のノードが縦方向で重ならないよう，スピーチの順番に沿って左右交互に配置した。これによりスピーチの境界が明確になり，反論のエッジが上方向にどの程度遡っているかも直感的に確認できるようになった。

Order. *Order* は反論の交差数の逆数を表す特徴量である。当初のデザインではすべての反論を薄い赤色で表示していたが，特にエッジが密集した際の視認性が低く，反論の交差が味方側のものか相手側のかを識別することが困難であった。そこで，肯定側の反論を赤色，否定側の反論を青色で表示し，線をより明瞭にすることで，各ディベータの反論における交差を容易に識別できるようになった。またそれに伴い，ノードも肯

定側を赤，否定側を青とすることで，各発言がどのチームから提示されたものか直感的に理解可能とした．また，POIが行われている際はPOIを受けている側のノードの色を薄く表示することで，相手側の発言中でもどちらのターンかがわかるように表現した．

Rally. *Rally* は再反論の連鎖の深さを表す特徴量であり，3章において議論の噛み合い度合いとの相関が最も高いことが確認された．ラリーの深さを視覚的に追跡できるようにするため，肯定側のADUを左側に，否定側のADUを右側に配置した．また，ノードを長方形とし，エッジの端点がノードの横の辺に来るよう設計した．これにより，反論のエッジが左右を行き来する様子からラリーの往復を容易に把握できるようになった．また，POIが行われている際にノードを相手側に表示することで，質疑への反論を通じたラリーの深まりも可視化されるようにした．

Interval. *Interval* は，同じ発言への反論がどの程度散在しているかを表す特徴量である．当初のデザインでは，発言の量に応じてノードの高さを変え，ノード内にテキストを表示していたが，この方式ではADU間の間隔の把握が困難であった．より直感的に同じターゲットへの反論の間隔を視覚的に理解できるよう，各ADUノードの高さを一定に設定し，テキスト表示を省略した．

以上の方針に基づき，3章で議論の噛み合い度合いとの相関が確認されたものを中心に，反論の構造的特徴を視覚的に把握できる可視化デザインを実現した．

4.3 DebaTube

4.3.1 必要要件

競技ディベートの動画を検索する際に，論題や大会名だけでなく，試合内容を手がかりに目的の動画を探したいという場合がある．この要求を満たすには，まず議論の全体像を俯瞰的に把握できる機能が必要である．動画の基本情報と紐付けて議論の流れを提示し，対立の構図や発言の関係を一目で把握可能とすれば，効率的かつ効果的な試合動画探索を支援できる．

次に，各発言および反論が肯定側と否定側のどちらに，どの順番で提示されたかを把握しやすくする必要がある．各発言の順序やチームの識別は，反論のバランスやチーム内の連携といった試合の特徴を見極めるうえで重要であるためである．

さらに，異なる試合の議論の特徴を比較しやすくする必要がある．関心のある試合を複数同時に見比べることができれば，その特徴の違いを把握しやすくなるためである．

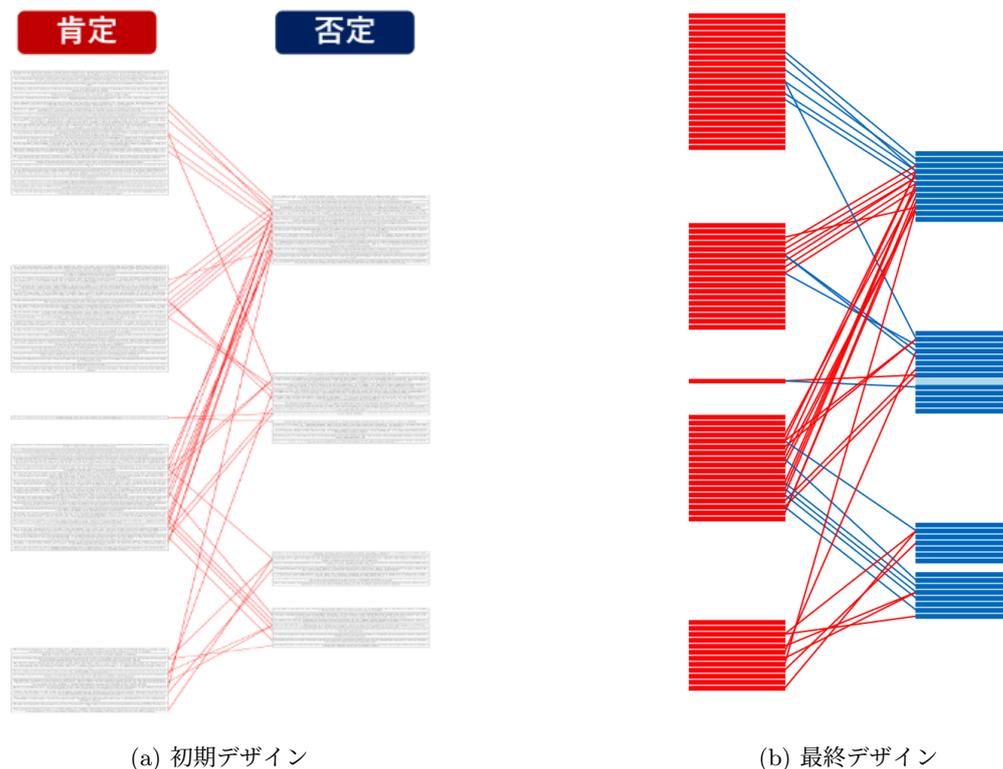


図 5: 反論グラフの初期デザインと最終デザインの比較

最後に、各発言や反論が提示された動画の箇所にアクセス可能とすることが必要である。動画探索を支援する上で判断材料が動画のどの箇所に紐づいているのかを把握可能とするため、議論全体の特徴を提示するだけでなく、実際に元となる動画のシーンを参照できるようにすることが求められる。

以上の要件をまとめると以下のようなになる。

- 要件 1: 試合の情報と結びつく形で、ディベートの構造を俯瞰的に把握可能とする
- 要件 2: 各発言とスピーチの関係や反論関係などのディベートの構造を直感的に把握可能とする
- 要件 3: 気になった試合をピックアップし、近くに再配置できるようにする
- 要件 4: 各発言や反論がある動画の箇所を直接参照できるようにする

そこで本研究では、これらすべての要件を満たす動画探索システム DebaTube²を実現する。

²<https://debatube.nkmr.io/ja>, <https://github.com/Non2d/DebaTube>

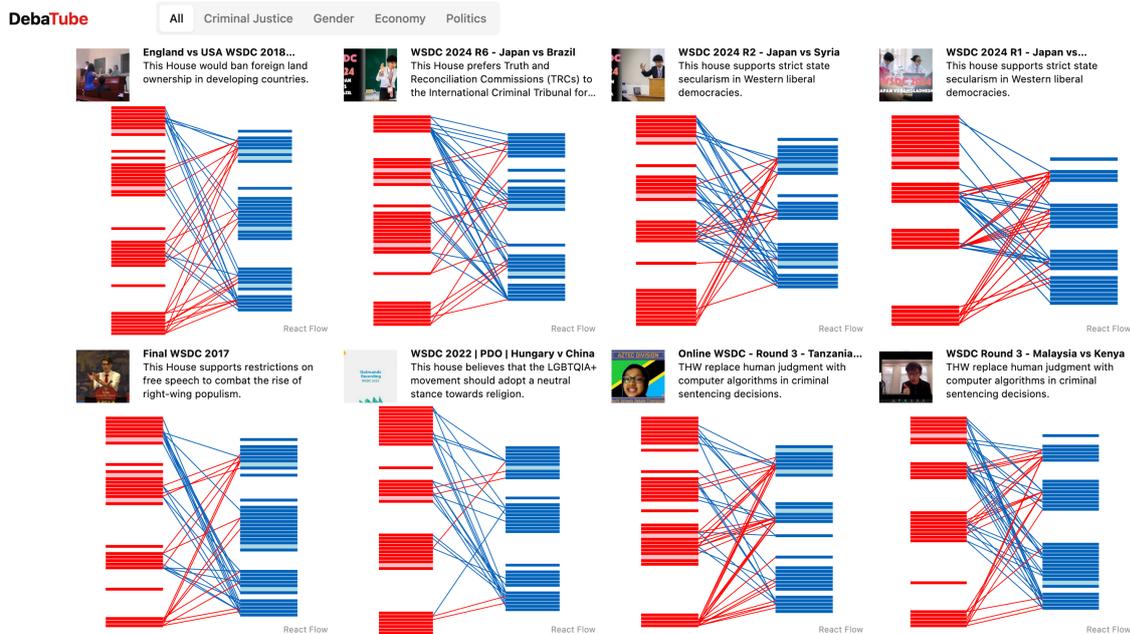


図 6: DebaTube のユーザインタフェース

4.3.2 実装とシステム概要

図6はシステムのUIの概観である。まず要件1および2を満たすため、タイトル、サムネイルと論題を表示するコンポーネントと、それに紐づく形で反論構造の可視化グラフを提示するUIを実装した。このグラフは、4.2節で定義した反論構造の可視化デザインに基づいている。さらに、要件3を満たすため試合のピン留め機能を、要件4を満たすためピン留めされた試合のノードをクリックすることでそのノードの発言が行われた動画の箇所を再生する機能を実装した。また、視聴する動画群のカテゴリを統制するため、選択したカテゴリの動画のみ表示する機能を実装した。

DebaTubeは、フロントエンドにTypeScriptとNext.js、グラフ描画のライブラリにReact Flow、バックエンドにPythonとFastAPI、データベースにMySQLを用いてWebアプリとして実装した。また、音声認識モデルはOpenAI社のwhisper-large-v3、LLMは同社のgpt-4o-2024-11-20、話者分離モデルはpyannote.audioを使用した。

4.4 予備調査

本章では、「試合での反論構造の可視化は動画探索においてどのような役割を担うか」という問いについて調査するため、ディベート経験者を対象にシステムを利用するタスクを行う。具体的には、DebaTubeを利用した動画探索において、ディベーターの行動が異

表 3: タスクで提示されるシチュエーション

番号	内容
練習	後輩がブレイン（話す内容や戦術を決めるリーダー的な役割）を担当する際、味方にどう役割分担を指示するか悩んでいます。
本番 1	後輩が、味方と似た話を繰り返してしまい上手く深掘りできないと悩んでいます。
本番 2	後輩が、一貫性のある主張を行う方法について悩んでいます。
本番 3	後輩が、最初のスピーチ（注：ターンのこと）に集中して反論することが多いが、このままで問題ないか悩んでいます。
本番 4	後輩が Opposition Whip（注：否定側 3 番手）を練習しており、議論を比較する方法について悩んでいます。

なると考えられる 1 つの動画を選定するタスクと良し悪しを見比べながら 2 つの動画を選定するタスクを設計し、可視化がディベータの動画探索にどう役立てられるかを調査する。

4.4.1 タスクの概要

練習タスクは各機能やタスクの流れなどのシステムの使い方に慣れてもらうために設けた。本番タスクは制限時間を 10 分とし、制限時間を超えた場合システムの使用をやめ、15 分経過するまでにアンケートへの回答を完了するよう指示した。また制限時間内であっても、タスクを完了できた場合は任意のタイミングで終了できるようにした。

各タスクは、与えられたシチュエーションの中で、指定されたカテゴリの 20 試合の動画の中から、質問に対する回答として最も適切な試合動画を、提案システムを使用した上で選ぶというものである。シチュエーションの一覧を表 3 に、質問の一覧を表 4 にそれぞれ示す。練習 1 は練習タスク、本番 1 から 4 は本番タスクで使用した。また質問に関しては、練習タスクでは全ての参加者に対し質問 1 を提示したが、本番タスクでは参加者ごとに質問を提示する順序を交互に入れ替えた。これは、参加者の動画探索行動やアンケート結果に対して、状況と質問が交絡することを避け、独立した要因として分析できるようにするためである。また、各タスクの最後にアンケートで「なぜその動画を選択しましたか？理由を教えてください」という質問と、「他に選ぶか悩んだ候補があった場合、それを選ばなかった理由をお答え下さい」という質問に回答してもらった。タスクはオンライン上で実施し、各参加者が所有している PC で行った。提案システムを活用して練習用タスク 1 つと本番用タスク 4 つに取り組んだ。

表 4: タスクで提示される質問

番号	内容
質問 1	後輩にどの動画を視聴して参考にするよう勧めますか？ 良し悪しを比較できる、良い試合の動画と悪い試合の動画を1つずつ選び、タイトル左の赤い番号でお答えください。
質問 2	後輩にどの動画を視聴して参考にするよう勧めますか？ 特に良いと思われる試合の動画を1つ選び、タイトル左の赤い番号でお答えください。

4.4.2 タスク用システム

システムは予備調査タスクのためにいくつか調整を行った。まず、分析のためにピン留めや動画再生といった各種操作の記録をタイムスタンプを含めてログとして取得する機能を実装した。また、各タスクが完了したタイミングを明確にするため、タスク完了ボタンを追加し、タスクが終了するたびにボタンを押してもらうようにした。さらに、タスクの回答で参加者が言及している動画を明確にするため、タイトルの左に試合固有の動画 ID を表示し、それを使って回答を記述してもらうようにした。

4.5 結果

実験参加者は半年以上の競技ディベート経験がある者 5 名であった。表 5 は参加者ごとの経験年数、所要時間、完了タスク数、動画再生数である。ここで、本タスクでは 2 つの参考になる試合を選ぶタスクを選択するタスクにおいて、多数の動画を見比べながら探索を行うことを想定していた。しかしながら、オンラインでの実験の実施や動画探索タスクの難易度が一部の参加者にとって想定以上に高かったことから、十分に指示を伝えられず、実験が成立しなかった参加者がみられた。したがって、1 タスクあたりの動画視聴回数が 4 回以下であった参加者 A・B・C の 3 名を、十分に参考になる動画について検討できていなかったと判断し除外し、以降の分析は残る 2 名を対象としたものを中心に行う。サンプルサイズが限られているため、本研究は探索的な予備調査として位置づけ、統計的な一般化ではなく、個別な事例の定性的な傾向の調査と今後の本格的な実験に向けた知見を得ることを目的とする。

表 6 は参加者 D と E の質問 1 および質問 2 のタスクそれぞれにおけるピン留め数、ピン留め解除数、ノードをクリックして動画を視聴した回数、視聴した動画の種類の数である。

この表から、両参加者とも質問 1 のタスクの方が質問 2 のタスクよりノードをクリックした回数が少なかったことが分かる。一方で、視聴した動画の種類の数やピンの操作の

表 5: 参加者ごとの経験年数およびタスクに関する統計量

参加者	A	B	C	D	E
経験年数	半年	1年	4年	4年	8年
完了タスク数	4	3	4	4	4
平均所要時間 (秒)	396.5	447.5	256.0	453.6	670.4
平均動画視聴数	3.3	3.3	0.3	8.3	9.8

表 6: 参加者 D, E の各質問における操作回数の統計量

	参加者 D		参加者 E	
	質問 1	質問 2	質問 1	質問 2
平均所要時間 (秒)	395.8	511.4	648.1	692.7
平均ピン留め数	5.0	4.5	8.5	8.0
平均ピン解除数	2.5	1.0	1.5	4.0
平均ノードクリック数	7.0	9.5	8.0	11.5
視聴した動画の種類	4.0	4.0	5.5	7.5

回数については両参加者間で共通する傾向はみられなかった。

また、全 8 タスク中 5 タスクで選ぶか迷った試合があるという回答が見受けられた。このことから、10 分程度の短時間の動画探索においても、参考になる動画を選択する上で複数の選択肢を吟味できていたと考えられる。

4.6 分析と考察

4.6.1 動画探索行動の統計量

前述したように、表 6 から、両参加者とも、タスクによらず時間をかけて動画探索に取り組む傾向がみられた。特に参加者 E は質問 1, 質問 2 の両方のタスクで制限時間一杯まで時間を費やしていることがわかる。また、視聴した動画の種類は選ぶ必要のある動画の種類が少ない質問 2 の方が多くなっている。さらに、参加者 D も選ぶべき動画数が少ない質問 2 において比較的長い探索時間を要している。これらのことから、選ぶ動画が少ないタスクにおいても時間をかけて幅広く動画を比較・検討する傾向が見受けられる。

ピン留め機能の使用回数に着目すると、参加者 E のピン留めおよびピン留め解除の使用数がかかなり多いことがわかる。特にピン留め機能は頻繁に使用され、選ぶ必要がある本数以上に動画をピックアップして検討する様子が観察された。一方、参加者 D もピン留め機能を積極的に使用しており、必要な動画の候補を見極めるために複数回のピン留

表 7: 参加者ごとの各タスクにおけるピン留めしてから
最初に視聴するまでの時間の平均 (秒)

	本番 1	本番 2	本番 3	本番 4
参加者 D	8.3	11.4	21.8	31.3
参加者 E	4.0	212.9	78.9	145.5

めと解除を行っている。

参加者 E において、質問 1 (良い試合と悪い試合を選ぶタスク) の方が質問 2 (良い試合だけを選ぶタスク) よりピン留めおよびピン解除が行われた回数が少ないことが分かり、タスクの初期段階から候補を絞って動画探索を進める傾向がみられた。一方参加者 D は、ピン留め数とピン解除数の差、つまり最終的にピン留めされたままの試合数は質問 1 の方が質問 2 より少なかった。これは、良い試合と悪い試合の両方を選ぶ質問 1 では、良い試合のみを選ぶ質問 2 と比べて最終的な候補が絞られやすいことを示唆している。しかし、探索の初期段階でのピン留め数自体には質問間で大きな差がみられなかったことから、参加者 D では候補の絞り込みが動画を選ぶ理由を固めるタスク後半の段階で行われていた可能性が考えられる。このように、タスクで提示される質問が動画探索に与える影響は、参加者 E では探索の初期段階から、参加者 D では探索の後半にみられるなど、参加者によって異なるタイミングで現れることが観察された。

動画再生までの時間からピン留め機能の活用方法に参加者間で大きな違いがみられた。表 7 は視聴された試合に関して、ピン留めされてから最初に視聴されるまでにかかった時間の平均である。この表から、参加者 D は、ピン留めした動画を視聴するまでの時間が長くとも 30 秒と短く、グラフの見比べをほとんど行っていない、あるいは見比べる場合でもピン留めをせず直接特徴を見ていたと考えられる。そのため、参加者 D はピン留め機能を比較検討のためのツールとして十分に活用していなかったと考えられる。一方、参加者 E は、2 番目以降のタスクではピン留めした動画の視聴までの時間が比較的長く、ピン留め機能を活用して気になる試合をピックアップし、グラフから特徴を見比べていたと考えられる。

なお、本分析の限界として、特に初心者のディベータにはディベートの特徴の理解が難しいことが挙げられる。表 5 にあるように、参加者 B は経験年数が 1 年であるが、本番 2 の議論の一貫性を学ぶ上で参考になる動画を選ぶタスクで、反論構造のどこに着目すれば良いかわからずタスクを完了できていない。このことから、今回設定したタスクの中には、反論構造の可視化と関係性を見出すことが難しい観点も含まれていた可能性がある。

る。今後はより適切な難易度のタスクの設計やシステムの対象とするディベータの経験年数の範囲を絞ること、あるいは可視化の情報量を増やすことが必要だと考えられる。

4.6.2 動画探索行動の時系列推移

試合動画を視聴した箇所を時系列に整理し分析した結果、両参加者ともいくつかのタスクで複数の試合間の同じスピーチに着目して動画を見比べる傾向がみられた。図 7 は参加者 D がタスク 4 に取り組んだ際の 250 秒経過時点、およびタスク終了時点でのピン留めの状態と視聴した動画の順序を示したものである。なお、操作のスタート地点に S、終了地点に G と表示している。このタスクでは否定側 3 番手の議論比較に関する質問が提示されており、参加者 D は良い動画として 73 番を選択している。図 7 から、参加者 D はまず肯定側 1 番手から 3 番手、および否定側の 2 番手と 3 番手を視聴し、それぞれの試合全体の流れを見ながら比較していた。これは、議論の比較を理解する上で、議論の流れを把握する必要があるためだと考えられる。このように、参加者 D は 73 番以外の試合についても、同じターンのシーンを複数の試合間で見比べる行動を行っていた。

同様の傾向は参加者 E でもみられた。図 8 は参加者 E がタスク 3 に取り組んだ際のピン留めの状態と視聴した動画の箇所を示したものである。この参加者は、まず 2 つの動画の肯定側 1 番手、2 人目のシーンを視聴したのち、ピン留めを解除した。その後、図 8 のように否定側 2 番手のシーンについて 5 つの試合を行き来しながら視聴していた。その中で、最終的に選択された 40 番は最もノードクリック回数が多く、4 回視聴されていた。また、途中で右上にある 38 番の動画の肯定側 1 番手を視聴しているが、1 秒以内に同じ試合の否定側 2 番手に移動したため、誤操作であったと考えられる。これらのことから、参加者 E は最終的に否定側 2 番手を中心に試合動画を見比べていたと考えられる。

また、参加者 E にはタスク 2 においても同様の傾向がみられた (図 9)。このタスクでは、一貫性のある主張を行う方法に関する質問が提示されており、この参加者は良い動画として 60 番を、悪い動画として 41 番を選択している。このタスクでは全ての動画のピン留めを最初に一斉に行った後、最後にピン留めした 41 番から順にピン留めした動画全ての肯定側の 1 番手や 2 番手を視聴している。これらの動画を選んだ理由として、参加者 E は「特に 1st speaker によってセットアップにおいて、論題の意味や現在の課題等が提示されている場合、一貫性が保持されやすいと判断 60 の動画では Gov 1st によりそれがなされているように感じ一方 41 では不十分に感じた」と回答しており、肯定側の冒頭のスピーチにターゲットを絞って見比べることで各試合のセットアップの状況を比較して

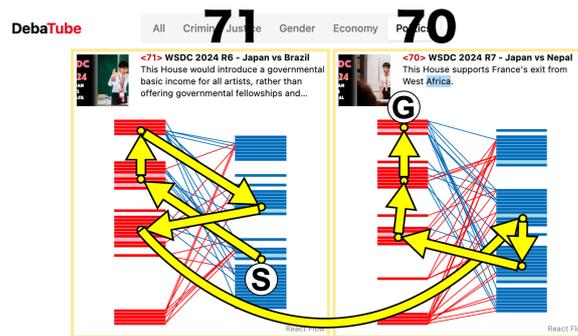


図 7: 参加者 D のタスク 4 の開始から 250 秒経過時点でのピン留めされた動画と視聴箇所への軌跡

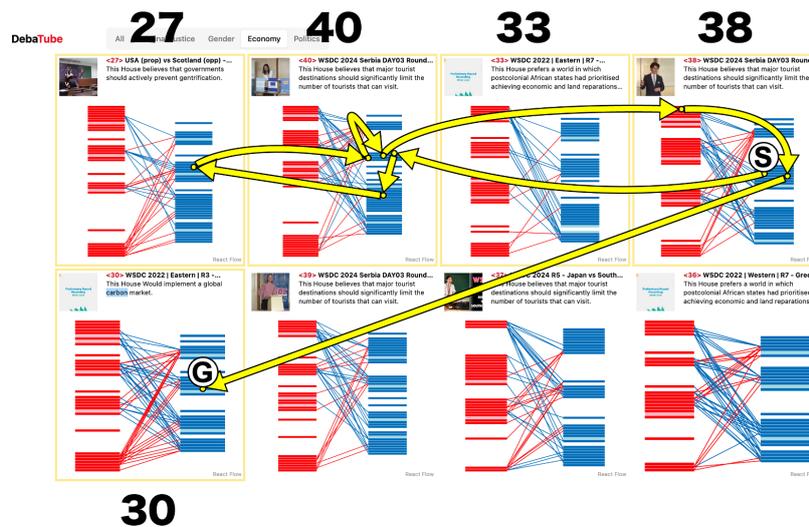


図 8: 参加者 E のタスク 3 の終了時点での UI とタスク開始 300 秒以降の視聴箇所への軌跡

いたと考えられる。

このように、否定側 3 番手などのように特定のターンに着目するよう指定されていなくても、複数の動画で同じターンを見比べる傾向が観察された。また、着目するターンを指定している場合であっても、参加者が自らの判断で注目する対象を広げ、異なる動画間で複数のターンに注目して探索を行う場合がみられた。

4.6.3 試合動画の選択理由

動画探索における反論構造の可視化の役割は、主に 2 通りみられた。1 つ目が、実際のグラフを見た上で動画の探索基準を決め、探索的に動画を探す手掛かりとして用いるパターンである。そして 2 つ目が、動画探索を始める前に、予め伝えたいメッセージと探すべき試合の特徴を決めた上で、それに合致する動画を探し、可視化はその動画が参考になる理由を説明する材料として用いるパターンである。

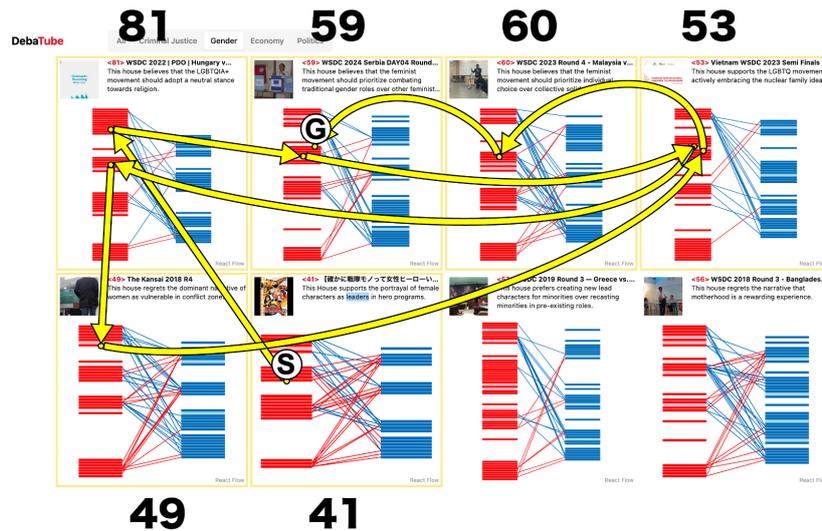


図 9: 参加者 E のタスク 2 の終了時点での UI と視聴箇所の軌跡

本実験では、前者のパターンが多く見られた。実際に、動画を選んだ理由の中で「エンゲージメント（注：反論）の対象のブロックに注目したとき、12の動画では複数のスピーカーが1つのmatter（注：論点）に繰り返しエンゲージしていることが見て取れ」や、「この動画の opp whip は反論以外の内容が多く、そうとなると多分比較をたくさんしていたのだろうと思ったから」といった、可視化を見てはじめてこういった特徴を持つ試合が参考になるかを判断していると思われる回答がみられた。

一方、後者の顕著なパターンとして、例えば主張の一貫性に関するタスクでは「特に 1st speaker によってセットアップにおいて、論題の意味や現在の課題等が提示されている場合、一貫性が保持されやすいと判断。60の動画では Gov 1st によりそれがなされているように感じ一方 41 では不十分に感じた」や、反論が最初に集中することの是非に関して、「原因はいくつか考えられるが、何であろうと後半のスピーカーが後半のスピーカーに適切に反論できる動画が参考になると判断」といったように、試合内容を見る前に予め参加者の中で着眼点や参考になるかの判断基準を決めていると思われる回答が複数みられた。

また、同じ動画を「参考になる良い動画」として選択していながら、異なる着眼点で動画探索を行っているという興味深い事例もみられた。図 10 は参加者 D, E の両者から良い試合として選択されたものである。この試合を選択した理由として、参加者 D は「比較軸を具体的に狭めることによって、前半までの抽象的な比較を抜け出して Voting Issue（勝敗を分ける争点）になりうる比較軸を出せていた」と、比較基準の明確さを評価していた。また、参加者 E は「相手の各スピーカーから出されていた内容に触れること、特に Gov Whip に触れられていることを評価。Opp Whip の役割として欠かすと勝敗に直結す

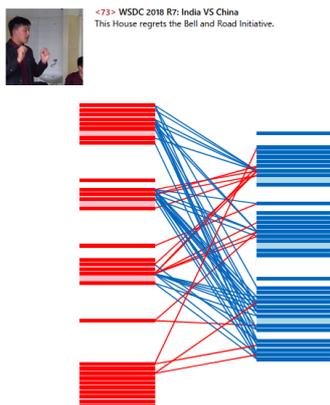


図 10: 両参加者に良い試合と評価された動画

る点であり、参考になると判断」と、相手側の発言に対し網羅的に反論できていることを評価していた。このように、同じ動画であっても試合の具体的な内容に踏み込んで評価するパターンの動画選択と、そこから一步引いた構造的な観点から評価するパターンの動画選択が、同じ結論に至ることがケースがみられた。

4.7 小括

本章では、競技ディベートにおける試合動画探索を支援するシステム DebaTube を実現し、その効果を調査するための評価実験を行った。提案システムは、試合の反論構造を可視化することで、ディベート経験者による試合の特徴の理解を支援することを目的としている。

まず、LLM を用いた反論構造の自動推定手法を構築し、YouTube 上の競技ディベート試合動画を対象にデータセットを構築した。その後、反論構造を可視化するシステムを実装し、競技ディベート経験者を対象に試合動画探索タスクを実施して、探索行動のログデータとアンケート結果を分析した。その結果、ピン留め操作やノードのクリック行動から、反論構造の可視化を活用しながら試合の特徴を把握する行動が見られた。

本予備調査は探索的なものであり、得られた知見は今後の本格的な検証を要するが、反論構造の可視化による学習支援の可能性を探る上での足がかりとなる知見が得られた。

第5章 DebaTube Liveの実装と利用実験による 検証

4章では、3章でデザインされた反論の構造的特徴を視覚的に把握できるシステムを活用して、さまざまな試合の特徴を概観しながら、参考になる動画を効率的に探索する行動が見られた。

ここで、ディベート学習では、動画視聴を通じた他者の事例から学ぶだけでなく、練習試合で自身が参加した試合の振り返りを通じた実践的な演習も盛んに行われている。次以降の試合に活かせるような有意義な振り返りを行うには、論題の分野における詳細な知識や強いロジックを知るだけでなく、反論のバランスや再反論といった、試合展開に応じた議論の戦略的および構造的な理解が重要である。実際に、教育学や専門職教育の分野では、実践後の振り返りが専門的スキルの発達に重要であることが広く認められている [75]。Kolb [76] の経験学習理論では、具体的経験、内省的観察、抽象的概念化、能動的実験という4段階のサイクルを通じて学習が深まるとされており、内省的観察の段階が効果的な学習において重要な役割を果たす。スポーツ科学においては、ビデオフィードバックを用いた振り返りが選手の意思決定能力や技術的パフォーマンスを向上させることが複数の研究で実証されており [77,78]、「知のスポーツ」とも称される競技ディベートにおいても、試合後に自身のパフォーマンスを広い視野で振り返ることが重要だと考えられる。

しかし、限られた時間で複数の論点の深掘りや反論が並行して展開されるという競技の特性上、試合後の振り返りでは個々の論証や反論の内容に関心が偏りやすく、試合全体を俯瞰的に省みることは困難である。

この競技ディベートの試合後の振り返りにおいても、反論構造の可視化は有効であると考えられる。可視化は複雑なプロセスの振り返りにおいても深い理解を促し学習効果を向上させることが明らかにされている [79,80]。また、4章で述べた通り、議論構造の可視化は議論の特徴理解を支援できる可能性が示唆されているが、振り返りにおいても迅速かつ俯瞰的な議論特徴の理解が重要である。以上のことから、反論構造の可視化には試合の振り返りにおいてもディベート学習を支援する効果が期待できる。

そこで本章では、4 章で設計した反論構造の可視化を改善したうえで自身の試合の振り返りに適用し、その効果を検証した。具体的には、まず「DebaTube」で実装した LLM を用いた反論構造の自動推定手法の精度および実行時間を大幅に改善し、試合の録音と紐付けて提示することで、俯瞰的な視点からの試合の振り返りを実現するシステム「DebaTube Live」を実装した。本システムは、試合中に使用するタイマー機能と録音機能を統合してスピーチごとに試合を録音し、試合終了後に録音音声から自動的に発言を抽出して発言間の反論関係をグラフとして可視化する。グラフ上の発言ノードをクリックすることで該当箇所の録音を直接再生でき、試合全体の議論構造を俯瞰しつつ気になる箇所の詳細を即座に確認できる。

次に、「自分が参加した試合の反論構造の可視化は、反論のバランスや順序といった反論の構造的特徴に着目した振り返りを促すか」という問いを検証するためにシステムの利用実験を実施した。具体的には、競技ディベートの練習試合を対象にシステムを利用して振り返るタスクを行ったのち、システムの効果に関するアンケートを行った。本章では、反論構造の自動推定手法の改善、システムの設計と実装、およびディベート経験者によるシステムの利用実験の結果について述べる。

5.1 LLM を用いた反論構造の自動推定手法の改善

これまでの反論構造の推定手法では、精度と処理時間の両面で複数の課題が存在した。本節では、その詳細とどのように解決したかについて述べる。

まず精度に関して、技術的課題とプロンプト設計における競技ディベートに関するドメイン知識の欠落という 2 つの観点から問題があった。技術的課題として、まず音声認識モデルが定義した、音声的な特徴に基づくセグメント単位での処理では区切りが不自然になることがあり、ADU が不自然に長くなる場合があった。また、ADU を主張と理由づけとして分類していたが、これにより ADU が過剰に細分化されてしまい、適切な反論判定に必要な前提となる文脈が失われ反論か否かの判断が人間にとってさえ困難なものとなっていた。さらに、ADU に ID を付与して反論判定を行う従来の手法では、LLM が判定に必要な情報を読み取りにくくなり、判定精度が低下していたと考えられる。ここで、LLM の性能を最大限に引き出すためには、自然言語コンテキストを直接処理させることが最も効果的であることが複数の研究で示されている [81–84]。そこで、反論ペア判定時に ID ではなくテキストを出力とするプロンプトおよびデータフローを再設計し採用した。加えて、GPT モデルはコンテキストウィンドウの制約から長文のディベート分析

には不向きであり， Gemini モデルの方が高い性能を示したため， Gemini モデルを採用した。

また， 競技ディベートのドメイン知識を活用することで反論構造推定の精度を向上させた。この競技では， ある程度どのディベータにも共通する決まったスピーチの構成がある。具体的には， インTRODクシヨN， 定義， 独立した反論， 2〜3 個の論点からなる本題， といった流れである。しかし， 従来の手法ではこうしたディベートの特性が十分に反映されていなかった。本システムでは， この情報を反論構造推定のプロンプトに組み込むことで， 定義の部分を反論と誤って判定するといった明らかなミスを大幅に削減した。

処理時間に関しては， 従来システムでは話者分離， ADU 判定， 反論判定などの処理が複数のパイプラインに分かれており， 試合終了後の即時振り返りには非現実的な処理時間を要していた。特に話者分離処理が計算コストの大きなボトルネックとなっていた。そこで， 話者分離の処理を後述するシステムデザイン上の工夫を凝らすことで自動化し， 精度を保ちながら限られた時間で完了できるようにした。

また， 文字起こしや反論判定などの自然言語処理にも時間が掛かっていたため， 話者ごとに分離されたスピーチごとのデータに対して並行して反論判定を行うことで処理時間を削減した。さらに， これまで分散していたすべてのパイプラインを単一の API に集約した。

5.2 DebaTube Live

本章では， 反論構造の可視化を基に試合後の俯瞰的な振り返りを支援するシステム「DebaTube Live」を実現した。本節では， その詳細について述べる。

5.2.1 必要要件

競技ディベートの試合の俯瞰的な振り返りを支援するためのシステムを実現するにあたり， 以下の 4 つの要件を設定した。

- 要件 1: ディベータが試合中に録音する手間を少なくすること
- 要件 2: 試合後に短時間で反論構造を推定できること
- 要件 3: 反論構造の推定や可視化を手軽に行えること
- 要件 4: 可視化グラフの各ノードが対応する録音箇所を直接参照できること

要件 1~3 を設定した理由は、システムを手軽に利用できることと試合直後の記憶が鮮明なうちに振り返れることが、可視化を効果的に活用してもらううえで重要なためである。また、要件 4 を設定した理由は、反論構造のグラフの各ノードがどの発言に紐づいているのかを即座に把握可能とすることで、反論構造を俯瞰しつつ、必要に応じて詳細な発言内容が確認でき、システムの実用性を高められるためである。

5.2.2 実装

5.2.1 節で述べたシステムの必要要件に基づき、反論構造の可視化から試合の振り返りを支援する機能を実装した。

まず要件 1 および 2 を満たすため、ディベータが普段からタイマーで時間を測りながらスピーチを行うことに着目し、各ディベータが自分のターンの開始と終了のタイミングで行うタイマー操作を利用することで、ターンごとに録音が行われるようにした。また、従来処理時間のボトルネックとなっていた音声ターンごとに分割する処理が不要となったため、録音データから音声認識モデルを用いた文字起こし、LLM を用いたスピーチごとの反論構造の推定、反論の可視化を行う工程が短時間で完了する実用的な処理速度が実現でき要件 2 が満たされた。次に、この一連の工程をすべて一つの機能に集約してボタン一つで全ての処理が完了するようにして、要件 3 が満たされた。最後に要件 4 を満たすため、グラフの各ノードをクリックすることにより対応する箇所の録音が再生される機能を実装した。

本システムはフロントエンドに TypeScript と Next.js, グラフ描画に React Flow, バックエンドに Python と FastAPI を用いて実装した。音声認識モデルとして OpenAI 社の whisper-1, 反論構造の推定に用いる LLM として Google 社の gemini-2.5-pro を使用した。

5.2.3 利用方法

図 11 および 12 はシステムのユーザインタフェースの概観である。ユーザは録音画面 (図 11) で各スピーチの録音の開始・終了を操作し、試合終了後に「Generate Graph」ボタンを押すことで反論構造の可視化を生成することができる。生成されたグラフは可視化画面 (図 12) で表示され、各ノードをクリックすることで音声プレイヤーが該当箇所にジャンプし、発言を直接確認することが可能である。なお、図 12 で「※ POI」と表示されているノードのように相手側のスピーチの中で孤立しているノードは、POI (Point of Information) と呼ばれる相手側のスピーチ中に割り込んで行われた質疑を表す。POI を

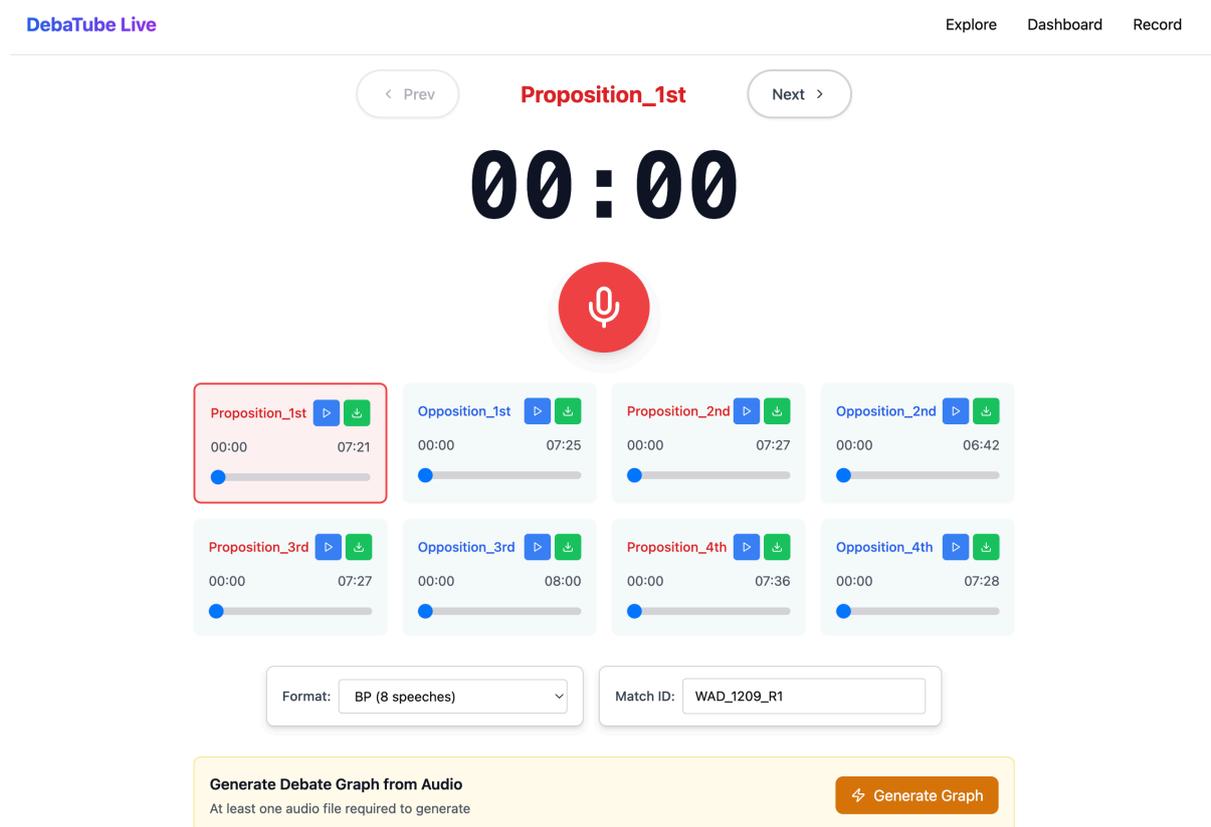


図 11: DebaTube Live の録音画面

このように配置した理由は、これまでの研究 [85] で POI のノードも含めて肯定側を左、否定側を右に配置したところ、どこからどこまでの ADU が何番目のスピーチか、直感的に把握しづらいという意見を受けたためである。POI は主に質問を目的としているため、反論先または反論元として判定されないようにした。

5.3 実験

本研究では、「自分が参加した試合の反論構造の可視化は反論のバランスや順序といった反論の構造的特徴に着目した振り返りを促すか」という問いを検証するため、システムを用いて振り返りを行う実験を実施し、その会話内容の傾向の違いを調査した。また、システムが主観的な振り返りやすさに与える影響について評価を行った。

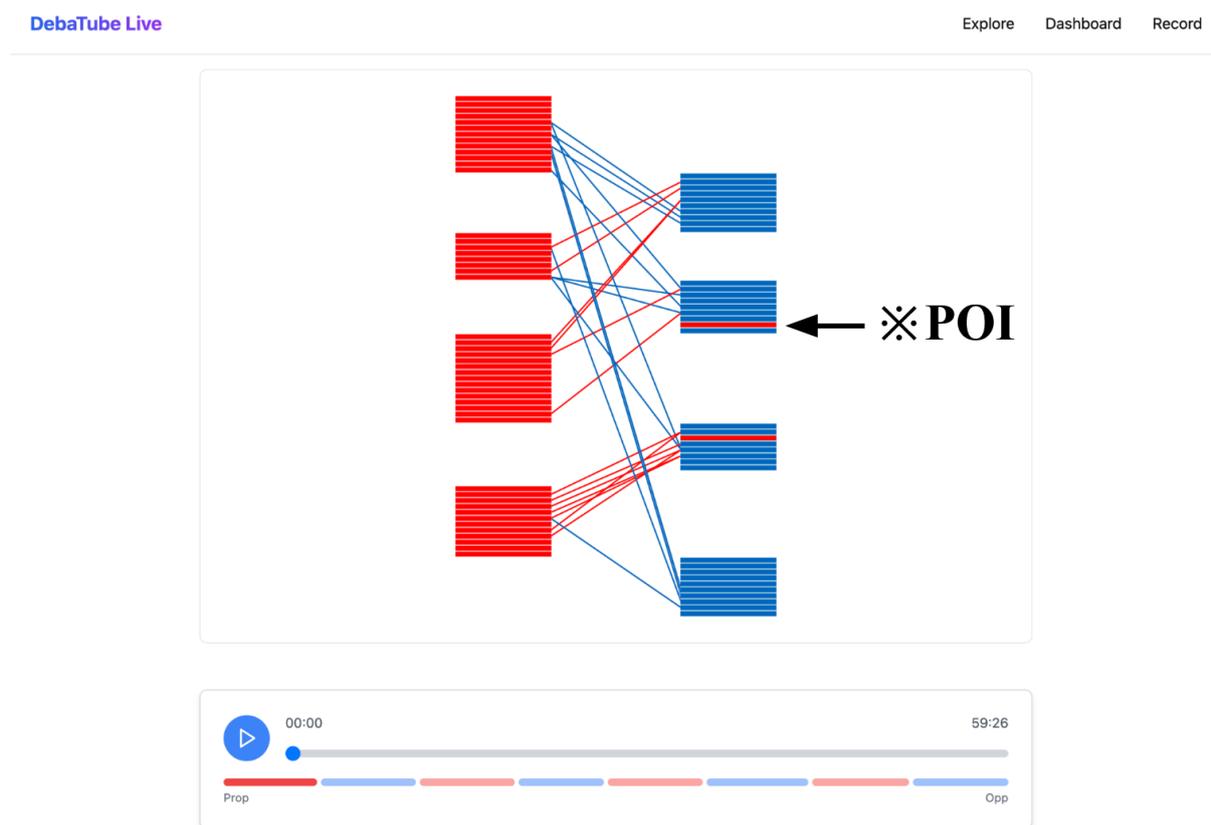


図 12: DebaTube Live の反論構造の可視化画面

5.4 British Parliamentary Style の概要

本章では、「British Parliamentary Style (以下BPスタイルと呼ぶ)」と呼ばれる、4チーム制の形式の試合を対象とした。この形式は他のフォーマットとは大きく異なる特殊なルールを採用しているため、本節では本フォーマットの概要について説明する。

BPスタイルは、世界大学ディベート選手権(WUDC)をはじめ、現在世界的に最も広く採用されているフォーマットである。本形式の最大の特徴は、1つの試合で4つのチームが同時に競う点にある。具体的には、肯定側(Government)と否定側(Opposition)は、それぞれ前半(Opening)と後半(Closing)に分かれ、反対の立場のチームとはもちろん、肯定側同士、否定側同士でも議論の説得力を競う。図13にBPスタイルにおけるチーム名、スピーチの順序と各ポジションの名称を示す。

各チームは2名のスピーカで構成され、持ち時間は1名あたり7分である。競技の結果は、各チームの議論の深掘りや効果的な反論などの貢献の度合いに基づき1位から4位まで順位付される。

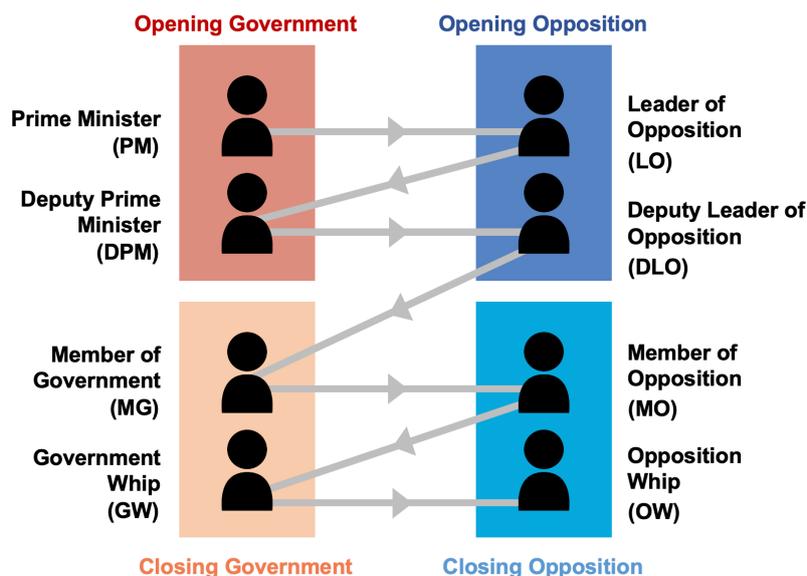


図 13: BP スタイルのチーム名およびポジション名

この形式では、原則として、後半チームは同じ立場をとる前半チームが提示した定義やスタンスを引き継がなければならない。前半チームと後半チームの間でも議論の一貫性が求められる。また、後半チームのチームが前半チームの同じ側のチームに対して反論を行うことは、ルール上禁止されていないものの推奨されない。したがって、後半チームには前半チームの議論と整合性を保ちつつ独自の貢献を行うことが要求される。

5.4.1 実験設計

本実験の目的は、反論構造の可視化を用いた場合に可視化を用いない場合と比べて、反論構造の着目が促されるかを検証するため、振り返りの傾向に違いが見られるか調査することである。そのためにまず、実験に参加するディベータを反論構造の可視化を使用して振り返りを行うチーム（提案システム群）と可視化を使用せずに振り返りを行うチーム（ベースライン群）に分けた。そのうえで、実際に試合後にチーム内で振り返りを行い、会話の中で話される内容の傾向を評価・比較した。また、可視化が振り返りのしやすさに影響するかを調べるため、アンケートによる主観評価を行った。

実験はBPスタイルのディベートの練習試合6試合を対象とし、前半の2チームを実験対象から除外した。これは、前半のチームの反論の機会は原理的に限られているため、振り返りにおいて反論構造に着目する必要性が少なく実験が成立しにくいと考えられたためである。

実験設計にあたり、以下の点に配慮した。

- 実際のディベート環境での実用性を調査するため、システムの使用やアンケートへの回答を除き、普段と同様の流れで実験が実施できるようにすること。
- 可視化システムは議論の良し悪しを判断するためのものではなく、第三者の評価に基づいて良い点や改善点を構造的に解釈するための補助ツールであるという位置づけを踏まえること。
- 可視化の解釈や議論の良し悪しの判断をジャッジに委ねてしまうことで、ディベータ自身が考えなくなる可能性を排除すること。
- 改善点の反省や良かった点を見つけるなど、幅広いユースケースを対象に可視化システムの効果を検証できるようにすること。

本実験は、大学のディベートサークルで行われた BP スタイルの練習試合を対象に実施した。試合のディベータは 1 試合あたり 4 チーム、延べ 24 チーム 48 名のディベータであった。このうち、実験には各試合 2 チーム、延べ 12 チーム 24 名のディベータが参加した。なお、重複して実験を行ったディベータもいたが、その場合必ず提案システム群とベースライン群の両方に同じ回数参加するようにした。また、チームメンバーは必ずしも固定されていない。

5.4.2 実験システム

提案システム群が振り返りで使用するシステムは、5.2 章で述べた DebaTube Live の可視化画面 (図 12) である。また、ベースライン群が振り返りで使用するシステムは、可視化画面から反論構造のグラフを除外し、音声プレイヤーのみを表示している。これにより、可視化の有無が振り返りに与える影響を比較できるようにした。

両システムともに、分析のためノードのクリックや音声プレイヤーのシークバーの操作などをタイムスタンプ付きのログとして記録した。

5.4.3 実験手順

5.4.1 節で述べた実験設計に基づき、以下の手順で実験を実施した。

1. BP スタイルの練習試合を実施し、試合全体を録音する。
2. 試合終了後、ジャッジが講評を行い、各チームの良かった点と悪かった点を指摘する。なお、統制のため必ず勝敗を判定する。

3. 後半チームの 2 チームを提案システム群とベースライン群に振り分け、チームごとに振り返りを行う。いずれの群においても振り返りは約 10 分間とし、その間はジャッジとは話さないこととした。また、両群ともに今回の試合で良かった点と悪かった点の両方について振り返るよう指示した。振り返りの際、両群とも紙などでメモした内容も参照してよいこととした。振り返りの様子は分析のため許可を取ったうえで全て録音した。なお、初めて提案システムを使用するディベータには、振り返りの前に筆頭著者がシステムの使い方や機能をレクチャした。
4. ディベータ全員に対してシステムおよび振り返りのしやすさを評価するアンケート（詳細は後述）を実施する。

また、実験の最後にアンケートを実施した。表 8 はアンケートの項目である。Q1～Q7 は「1: 難しかった」から「5: 簡単だった」、Q8～Q10 は「1: そう思わない」から「5: そう思う」の 5 段階リッカート尺度で回答を求めた。いずれの項目も、スコアが高いほど肯定的な評価を示す。なお、ディベータの多くが非日本語話者であることを考慮し、実際のアンケートには英訳を併記した。Q1～8 はシステム使用により主観的な振り返りの容易性が向上するかを調査するために全員が回答し、Q9 および 10 は提案システムにおける反論判定の妥当性や可視化の有用性について評価するために提案システム群のみが回答した。

5.4.4 振り返り内容の分析手順

可視化システムが反論構造に着目した振り返りを促すか調査するためには、実際の発言内容を体系的に分類する必要がある。そのため、本研究では振り返りの文字起こしに対して、定量的な発言の傾向の評価及び比較を可能とするコーディング分析を採用した。

具体的には、まず録音された振り返りの音声データから作成された文字起こしを発言ユニットに分割した。発言ユニットの区切りは話題の変化を基準とし、新しいトピックに移行した時点で新たなユニットを開始するものとした。また、分割された発言ユニットに対してカテゴリの分類を行った。カテゴリは、反論構造の特徴に関する 5 つのカテゴリを事前に定義したうえで、それ以外の発言については探索的にカテゴリを定義する方針をとった。事前に定義したカテゴリは以下の 5 つである。

- 誰が誰に反論していたかへの言及
- 反論の集中度合い/バランスへの言及

表 8: 振り返りの容易性とシステム評価（提案システム群のみ）のアンケート項目

	質問項目
Q1	自分が誰に対して反論していたかを思い出すことはどの程度簡単でしたか？
Q2	自分のチームと相手のチームの主張の関係を思い出すことはどの程度簡単でしたか？
Q3	振り返るにあたり、自分のチームと相手のチームの主張の関係を整理することはどの程度簡単でしたか？
Q4	うまくいかなかったところや改善すべき点を見つけることはどの程度簡単でしたか？
Q5	その改善点を、次以降の試合でどう直すか具体的に考えることはどの程度簡単でしたか？
Q6	うまくいったところを見つけることはどの程度簡単でしたか？
Q7	そのうまくいったところを、次以降の試合でも続ける方法を考えることはどの程度簡単でしたか？
Q8	今回の振り返りは、別の論題の試合でも活かせそうだと感じましたか？
Q9 (提案システム群のみ)	反論のグラフは振り返りに役立てやすかったですか？
Q10(提案システム群のみ)	反論判定は妥当だと感じましたか？

- 再反論に関する言及
- 反論の順序への言及
- 反論の多さへの言及

これら以外の発言は一度「その他」として分類し、その後データを順次分析しながら新たなカテゴリが出現しなくなるまで探索的にカテゴリを追加していった。なお、発言ユニットの分割とカテゴリの分類はすべて筆頭著者が手作業で行った。

5.5 結果

5.5.1 システムの動作

本実験で使用した 6 試合の録音時間は平均 58.2 分 (SD=1.6 分, 範囲 56.2-60.4 分) であった。これらの録音データに対して、文字起こしから可視化までの処理時間を測定した結果、平均 7.4 分 (SD=0.3 分, 範囲 7.0-7.8 分) であり、実験の妨げにならなかった。

このことから、試合終了後すぐに可視化が生成される実用的な振り返り支援システムを実現できたといえる。

5.5.2 反論構造に着目した振り返りの傾向

まず、試合直後の反論構造の可視化が俯瞰的な振り返りを促すかを調査するために行った、振り返りの録音のコーディングの結果を述べる。

最初の 3 試合分のデータを分析した時点で、事前定義の 5 カテゴリに加え、以下の 5 つのカテゴリが特定された。

- 内容の改善点に関する言及
- テクニックなど内容以外の改善点に関する言及
- 良かった点に関する言及
- ユーザが得た気づきや学びへの言及
- システムや機能への言及

4 試合目以降は新たなカテゴリが出現しなかったため、これら計 10 のカテゴリが最終的なコーディングの分類として確定した。

表 9 に、ベースライン群と提案システム群それぞれの各カテゴリにおける発言ユニット数を示す。両群ともに合計 60 ユニットが記録された。反論構造の特徴に関するカテゴリについて見ると、反論構造に関する言及の合計は、提案システム群が 19 ユニット、ベースライン群が 2 ユニットであり、提案システム群で顕著に多かった。一方、「内容の改善点に関する言及」についてはベースライン群が 32 ユニット、提案システム群が 10 ユニットと、ベースライン群で顕著に観察された。また、「ユーザが得た気づきや学びへの言及」もベースライン群が 12 ユニット、提案システム群が 4 ユニットであった。「良かった点に関する言及」についても、ベースライン群が 8 ユニット、提案システム群が 3 ユニットとベースライン群の方が多かった。

5.5.3 振り返りのしやすさに関するアンケート結果

5.4.3 節で設定したアンケートの結果を述べる。なお前述のように、いずれの項目もスコアが高いほど肯定的な結果を示す。

表 9: 振り返り内容のコーディング分析

言及内容	ベースライン	提案システム
誰が誰に反論していたか	0	3
反論の集中度合い及びバランス	2	8
再反論	0	1
反論の順序	0	1
反論の多さ	0	6
内容の改善点	32	10
テクニックなど内容以外の改善点	6	10
良かった点	8	3
ユーザが得た気づきや学び	12	4
システムや機能	0	14
合計	60	60

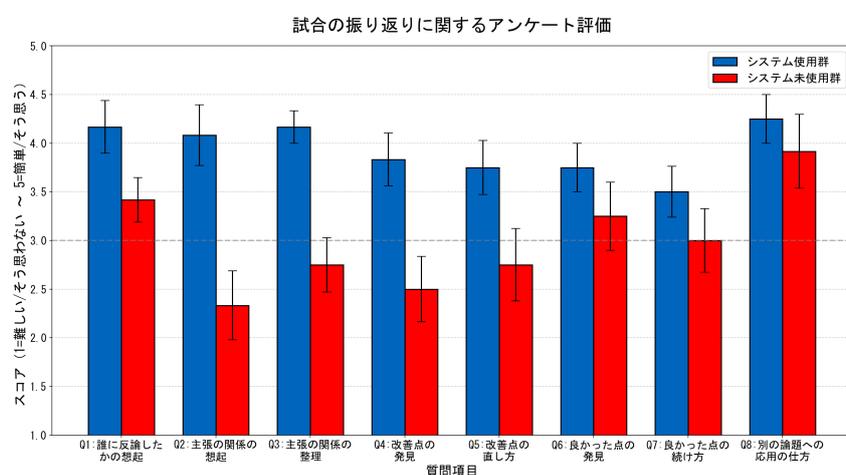


図 14: 振り返りの容易性に関する評価アンケートの結果

まず、図 14 は振り返りの容易性に関する Q1～Q8 における、各質問項目での提案システム群とベースライン群の平均値と標準誤差の結果である。このグラフから、すべての質問項目において提案システム群の方が高いスコアを示していることが分かる。特に、平均値に関して、Q2「主張の関係の想起」では両群間で 1.75 の差が見られ（提案システム群: 平均 4.08, ベースライン群: 平均 2.33）、最も大きな差を示した。また、Q3「主張の関係の整理」、Q4「改善点の発見」、Q5「改善点の直し方」で 1.00 以上、Q1「誰に反論したかの想起」で 0.75 の差が観察された。一方、Q6「良かった点の発見」、Q7「良かった点の続け方」Q8「別の論題への応用の仕方」では差は 0.50 以下にとどまった。

このグラフから、すべての質問項目において提案システム群の方が高いスコアを示していることが分かる。特に、平均値に関して、Q2「主張の関係の想起」では両群間で 1.75

の差が見られ（使用群: 平均 4.08, 未使用群: 平均 2.33）、最も大きな差を示した。また、Q3「主張の関係の整理」Q4「改善点の発見」Q5「改善点の直し方」で 1.00 以上、Q1「誰に反論したかの想起」で 0.75 の差が観察された。一方、Q6「良かった点の発見」Q7「良かった点の続け方」Q8「別の論題への応用の仕方」では差は 0.50 以下にとどまり、これらの項目ではシステムの効果は限定的であった可能性が示唆された。

次に、システムの機能の評価に関するアンケートの結果を述べる。Q9「グラフは役に立てられたか」Q10「反論判定は妥当か」は「1: そう思わない」～「5: そう思う」の 5 段階評価であり、スコアが高いほど肯定的な評価を示す。まず、Q9 の振り返りへの有用性については、平均 4.60 (SD=0.70) と高く、90% が肯定的評価 (4-5) を示し、70% が最高評価 (5) であった。Q10 の反論判定の妥当性については、平均 3.70 (SD=1.34) であり、60% が肯定的評価 (4-5) を示した一方で、30% が否定的評価 (2) であった。ディベート経験 (1: 半年以下, 2: 半年～1 年, 3: 3 年以上) および大会戦績 (1: 大会に出場した経験がない, 2: 大会に出場したことはあるが予選を突破したことがない, 3: ルーキー部門でのみ予選突破したことがある, 4: 一般部門で予選突破したことがある) と各評価項目との相関を分析したところ、大会戦績と Q9 の振り返り有用性の間に中程度の負の相関が見られ、実力のある者ほど振り返りシステムが役に立ちにくいと評価する傾向が示唆された。一方、ディベート経験年数および大会の戦績と反論判定妥当性の評価には相関がほとんど見られなかった。

5.6 分析

前述の結果から、DebaTube Live は反論の構造的特徴に着目した振り返りを促進する可能性が明らかになった。また、反論構造の可視化が主観的な振り返りの容易性を向上させることが明らかになった。本節ではこれらの結果についてより詳細に分析する。

5.6.1 反論構造への着目の促進

振り返りのコーディング結果から、提案システム群では反論構造の特徴に関する言及が合計 19 ユニット観察されたのに対し、ベースライン群ではわずか 2 ユニットであった。両群の総発言ユニット数が同一 (各 60 ユニット) であることを踏まえると、この差異は提案システムが反論構造に着目した振り返りを効果的に促進したことを示している。興味深いことに、ベースライン群で観察された反論構造への言及は「反論の集中度合い/バランスへの言及」の 2 ユニットのみであった点である。これは、可視化なしでは反論構

造という観点自体がディベータに認知されにくいことを示唆している。一方、提案システム群では5つの反論構造カテゴリすべてにおいて言及が観察され、特に「反論の集中度合い/バランスへの言及」(8ユニット)と「反論の多さへの言及」(6ユニット)が多くみられた。これは、ノード間のエッジの分布やエッジの総数といったグラフの視覚的特徴が、ディベータの興味や注意を効果的に誘発したためと考えられる。

ベースライン群において、「内容の改善点に関する言及」が60ユニット中32ユニットと半数以上を占めていたことから、従来の振り返りでは、論題ごとの話題に特化した具体例や論理といったアイデアに関するものが大部分を占めていることが明らかになった。もちろんこうした知識ベースの振り返りも重要ではあるが、BPスタイルでは、突拍子もない政策や、だれも知識を持っているはずがないその場での即興力が試される論題が出題されることもあり、幅広い領域に対応しなければならない。こうしたことを踏まえると、現状の振り返りは特定の分野の知識に特化しすぎている可能性があると言える。

一方で、提案システム群では「内容の改善点に関する言及」(10ユニット vs 32ユニット)、「良かった点に関する言及」(3ユニット vs 8ユニット)、「ディベータが得た気づきや学びへの言及」(4ユニット vs 12ユニット)がいずれもベースライン群より少なかった。この結果は、限られた振り返り時間(約10分間)において、反論構造の探索に認知資源が割かれた結果、内容面の深い考察に充てる時間が相対的に減少した可能性を示唆している。また、「システムや機能への言及」が提案システム群でのみ14ユニット観察されたことから、システムの操作や理解に一定の時間がかかっていた可能性が窺える。

ただし、この傾向が振り返りの質の低下を意味するかどうかは慎重に解釈する必要がある。反論構造という新たな観点からの振り返りは、従来の内容中心の振り返りでは得られにくい種類の学びをもたらす可能性がある。実際、「テクニックなど内容以外の改善点に関する言及」は提案システム群でやや多く(10ユニット vs 6ユニット)、反論構造の可視化が戦略的・技術的な側面への気づきを促進した可能性が示唆される。

5.6.2 振り返りのしやすさの主観評価

アンケート結果から、提案システムは振り返りのしやすさを向上させる可能性が確認された。

Q2およびQ3において、提案システム群はベースライン群と比較してより肯定的な評価を示した。このことから、反論構造のグラフが、議論における主張間の関係性を外在化・可視化することで、認知的な想起と整理の負担を軽減したのと考えられる。

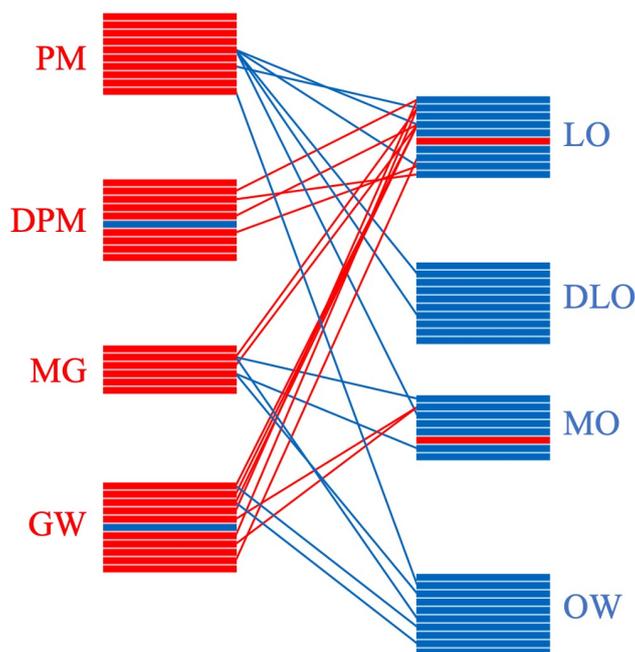


図 15: 試合 1 の反論構造の可視化. ポジション名と所属チーム名の略称が記載されている.

また, Q1 および Q4 についても同様に提案システム群の評価が高かった. グラフのノードとエッジの構造が反論の対象を明示的に示すことで, 記憶の手がかりとして機能したことを示唆している. Q4 の結果については, 反論構造の可視化により, 反論が不足している箇所や偏りが視覚的に把握しやすくなったことが寄与していると考えられる.

Q6, Q7, Q8 については両群間に明確な差は見られなかった. この結果は, 提案システムが主に改善点の発見や議論構造の理解を支援する一方で, 成功体験の特定やその汎化については追加の支援が必要である可能性を示唆している. うまく行った点の振り返りには, 単なる構造の可視化だけでなく, 何が良い反論構造であるかという規範的な情報の提示が必要とされている可能性がある.

5.7 考察

本節では, 特徴的な振り返りが行われた個別の事例を取り上げ, 提案システムの有無による振り返りの質的な違いについて詳細に考察する.

5.7.1 試合 1：本議会は、学校は歴史を偉大な人物の産物としてではなく、社会的および制度的要因の産物として教えるべきであると信じる

可視化あり条件（否定側後半チーム，2 位）では，ポジションによる戦略の最適化に関する気づきが顕著であった。参加者は「肯定側前半チームへの反論を控える」という戦略を意図的に実行しており，図 15 において否定側後半チーム（右下）から肯定側前半チーム（左上）への反論のエッジが 1 本のみであることを確認した上で，以下のように振り返った。

「今回，可視化から肯定側前半チームに対して 1 回しか反論していなかったと分かります。そして彼らの話を補強しました。なので，否定側前半チームとあまり被らないように議論できていたとわかります。」

また，システムが振り返りの網羅性を向上させる可能性についても言及があった。

「普段私は振り返りで全ての発言を取り上げるわけではありません。明らかに大きな塊（論点のまとまり）だけを取り上げます。でもこのシステムがあれば，見れば分かります。」

一方，ベースライン条件（肯定側後半チーム，4 位）では，具体例の追加や論点の拡張といった内容面での改善が中心であり，誰の論点に対して反論すべきだったかという構造的な視点は欠落していた。

5.7.2 試合 2：高齢化が進む社会において，本議会は，政府が国際的なマッチング（結婚紹介）機関に補助金の拠出を支持する

可視化あり条件（否定側後半チーム，2 位）では，図 16 の否定側後半チームにおいて，OW が肯定側 4 人全てに反論のエッジが伸びていることが確認され，チームメイトから「バランスの取れた反論」として評価された。

しかし OW は可視化の解釈において批判的な観点も示した。図 16 では以下のように述べ，反論の多さと勝敗の関係について批判的に分析した。

「(反論の) 線が多いから勝つわけでもないんだから。逆に肯定側前半チームはあんま線がないけど 1 位取ったから。」

この発言は可視化を反論が多いほど良いと示すためのものだと誤解していることが一因となっていると考えられる。しかし，本研究はこれまで述べてきたように，反論が多

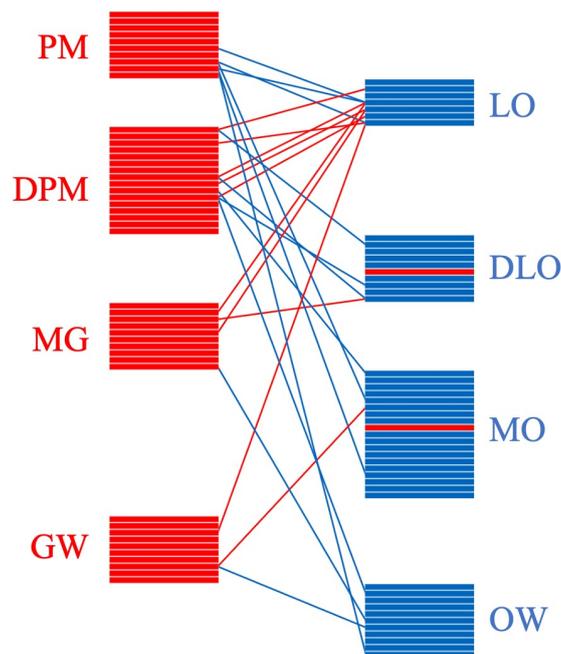


図 16: 試合 2 の反論構造の可視化

いほど良いと主張するものではなく、あくまで構造的観点から振り返ることが重要だとしている。この認識の齟齬はシステムの利用経験が浅いために生じたと考えられる。今後、長期的な実験により可視化への理解を深められれば、こうした課題は解決可能であると考えられる。

また、MOはチームメンバーにADU数が多く議論を多面的に展開できていたこと（図16参照）を称賛されたが、本人は「多分アサーション（理由づけのない新規の主張）だけ入ってる。（ADUのノードが）層になってるけど、実際の内容は一緒だと思う」と自己批判的な解釈を示した。これらは可視化を過信しない健全な批判的姿勢であり、可視化から主張をどのように展開べきかについてメタ的に理解を深められている可能性を示唆する。

また、MOは議論のADUへの分割に関して十分に妥当だと評価していた。このディベータは連続する3つのノードをクリックしたところ、それぞれのノードの冒頭部分が論点のナンバリングと一致していたことを受け、「おおー、なるほど。なんか、結構ちゃんと分けられてるね。」と評価していた。このことから、ADUの境界には論点のナンバリングの区分と一致していることが重要であること、本手法でそれが十分に反映できていることがわかる。

一方で、ベースライン条件（肯定側後半チーム、4位）では、論点の内容的な重複や不

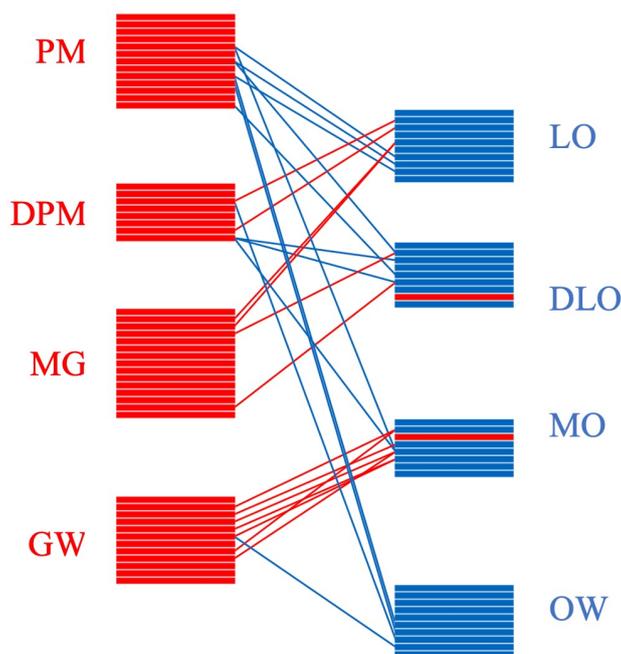


図 17: 試合 3 の反論構造の可視化

足に関する分析が中心であり、反論の定量的把握は見られなかった。

5.7.3 試合 3：実現可能性を考慮しなければ、本議会はウラジーミル・プーチンを暗殺する

本試合は可視化システムの効果が最も顕著に表れた事例である。

まず、可視化システムの興味深い効果として、参加者が自らの特定の ADU に対して関心を示す場面が観察された。

「MG: このノード好きだな。私が持ち出した新しいトピックだ GW: へー、本当？ MG: そう、長方形のひとつひとつが、スピーチで出した新しいトピックを表してる。」

この発言は、可視化が自らの貢献をノードとして提示することを通じて、議論への主体的な関与を促進することを示唆している。

また、図 17 の肯定側後半チームにおいて MG が 3 本の反論のエッジを出している様子を確認し、事前に立てた戦略の実行状況进行评估する場面も見られた。

「MG: うん、私に関してはかなり正確に出てると思う。主な反論は 3 つだけだから。(中略) 相手のケースを 3 つの弱いポイントに要約して、その弱いポイ

ントを論破するっていう。最後の反論はあんまり深掘りできなかったと思うけど。これについて話す時間がなかったから。」

ここからは、参加者が可視化を通じて「先輩の戦略を模倣する」という学習プロセスの成否を評価していることがわかる。

さらに最も顕著な効果として、時間配分の問題とその原因が可視化された点が挙げられる。図 17 では GW からの反論のエッジが多数出ているものの、そのほとんどが否定側後半チームに集中し、否定側前半チームへの反論が少ない。参加者はこの偏りを認識し、以下のように述べた：

「GW: うん。私は... これを見ると、反論はたくさんした。でももっと比較の中で反論を伝えることに集中すべきだったな。(中略) 特定した 3 つの指標についても話せなかった。そう、ライフスキル、安定性、長期的影響っていう 3 つを特定したんだけど、ライフスキルについて話す時間がなかった。」

「GW: ここ見るとわかるけど、反論しすぎたんだよね。反論に時間を浪費すぎて、自分たちの話の独自性を出せなかった。(中略) で、次は…」

さらに、MG も対戦相手へのエンゲージメント不足についても「システムを見て気づいたんだけど、否定側後半チームにしか反論してなかった。他へももっと反論すべきだったのに」と反省している。単に時間配分が誤っていたこと自体は試合中から認識していた可能性があるが、可視化を見ることで「反論に時間を割きすぎたこと」や「反論対象の構造的な偏り」といった具体的な原因を明確に特定できている点が可視化が無ければ気付けなかった可能性が考えられる。

一方で、ベースライン条件（否定側後半チーム、3 位）においても時間配分の問題への言及はあった：

「私はジャッジの方に聞いたんだけど、私のスピーチで完璧じゃなかったのは…時間がなくなっちゃったこと。(中略) 反論に時間を使いすぎた。」

しかし、この気づきはジャッジからのフィードバックによって得られたものであり、内容の改善点に関する発言が大半を占めていた。つまり、経験豊富な指導者からのフィードバックがあれば時間配分レベルの振り返りは可能であるといえる。しかし、特に経験の浅いディベータにとって、自分自身でそうした課題に気付くことは困難であると考えられる。

以上のことから、本システムは、参加者に論点に対する興味を抱かせ、戦略が思った通りに実行できているか状況を確認できるようにするだけでなく、時間配分の失敗原因や反論対象の偏りといった構造的な問題を明確にする、あるいは気づかせる効果を持つ可能性が明らかになった。特に、ベースライン条件では他者の指摘に依存していたような気づきを、可視化システムのディベータが自力で得られていることは、本システムが振り返りの質を高める上で効果的であることを示している。

5.7.4 小括

本章では、自分が参加した試合の反論構造を可視化するシステム「DebaTube Live」を提案し、実際の練習試合環境における利用実験を通じて、反論構造の可視化が振り返りに与える効果を検証した。具体的には、タイマー機能と録音機能を統合し、試合終了後に録音音声から自動的に発言を抽出して反論関係を推定・可視化するシステムを実装した。British Parliamentary スタイルの練習試合 6 試合を対象とした利用実験の結果、提案システムを使用した群では反論構造の特徴に関する言及が顕著に増加し、反論のバランスや集中度合いといった構造的観点からの振り返りが促進されることが確認された。また、主観評価の結果から、主張間の関係の想起や整理、改善点の発見といった側面において主観的な振り返りの容易性が有意に向上することが示された。

これらの取り組みを通じて、LLM を用いた反論構造の自動推定手法を大幅に改善し、約 1 時間の試合録音に対して文字起こしから可視化までの処理が短時間で完了する実用的なシステムを実現した。また、4 章では他者の試合を対象とした限定的な検証にとどまっていた反論構造の可視化の有用性について、本章では参加者が自身の試合の振り返りに可視化を活用できることを、実際の練習試合環境における利用実験を通じて実証的に確認した。特に、可視化を通じて時間配分の失敗の原因や反論対象の偏りといった構造的な問題を参加者が自力で発見できることが示され、反論構造の可視化がディベート学習における振り返りの質を高める上で有効であることが明らかになった。

しかし、本章の結論には制約が二点ある。第一に反論関係の推定精度に改善の余地があり、結果に影響を与えた可能性が排除できていない点である。5.7.2 節で述べたように、実験では議論の ADU への分割については高い評価を得た。しかし、5.5.3 節から分かるように、Q10 の結果から 30% のディベータが反論判定の妥当性に否定的な評価を示しているため、この工程における精度に課題があると考えられる。反論判定の納得性はディベータにシステムを積極的に活用してもらおううえで不可欠であるため、今後改善したう

えでシステムを再検証する必要がある。

第二に、長期利用時の効果が未検証である点である。本実験では各参加者の利用回数が最大でも 2 回であり、長期利用による影響は明らかでない。特に、アンケートでは大半のディベータが自分の声を聴くことへの忌避感を示しており、長期利用の障壁となりえる。ただし、この忌避感はシステムへの不慣れに起因する一時的なものである可能性があること、競技ディベートではジャッジをスピーチで魅了することが求められる [4] ため、自身の話し方を見直すことが重要であることから、自身の録音を振り返る仕組み自体が不適切だと断定するのは尚早である。したがって、現在のシステムデザインを維持したうえで、長期利用が忌避感や可視化の効果に与える影響を調査することが必要だと考えられる。

第6章 総合考察と展望

6.1 総合考察

本研究では、「競技ディベートにおいて俯瞰的な試合特徴の把握が困難である」という問題に着目し、反論構造の定量化手法および可視化によるディベート学習支援手法を提案した。本節では、これまでに行った研究を通じてその目標を達成できたかについて確認を行う。まず、3章では反論構造の定量的な特徴が議論の噛み合い度合いを示すかについて検証する実験を行った。4つの仮説に基づく特徴量（Distance, Interval, Order, Rally）を定義し、手作業で構築したコーパスを用いて専門家および LLM による評価との相関を分析した。その結果、重回帰分析による予測値と評価値の相関係数は 0.609 であり、中程度の正の相関が確認された。特徴量の相対的重要度（標準化係数の絶対値比率）は、Rally (40.7%) が最も高く、Distance (28.8%), Interval (25.3%) も中程度の寄与が確認された。これにより、反論構造の定量的特徴が議論の噛み合い度合いをある程度反映することが示唆され、勝敗判定から独立した議論を俯瞰的かつ客観的に評価する手法の基盤が確立できたと言える。

次に、4章では LLM を用いた反論構造の自動推定手法の基礎を構築し、これを用いて試合動画探索を支援するシステム「DebaTube」を実装した。評価実験の結果、ディベート経験者が反論構造の可視化を手がかりとして動画を比較検討し、様々なユースケースに沿って有用な構造的特徴を持つ試合を効率的に探索できることが確認された。特に、複数の動画の同じターンに着目して見比べる行動や、可視化を探索的な手がかりとして用いるパターンとストーリーテリングの材料として用いるパターンの両方が観察された。

最後に、5章では自分が参加した試合の振り返りを支援するシステム「DebaTube Live」を実装した。このシステムでは、第4章で課題となっていた推定精度と処理時間の問題を解決するため、ディベート特有のドメイン知識を反映したプロンプト設計、並列処理の導入などの改善を行い、試合終了後に短時間で反論構造を可視化できる実用的なシステムを実現した。British Parliamentary 形式の練習試合を対象とした利用実験の結果、提案システムを使用した群では反論構造の特徴に関する言及が顕著に増加し（19 ユニット

vs 2 ユニット), アンケートにおいても主張間の関係の想起や整理に関する項目で有意な改善が確認された。これにより, 反論構造の可視化が俯瞰的な振り返りを促進することが実証された。

以上の結果より, 本研究が提案した反論構造の定量化および可視化は, 定量的評価・動画探索・振り返りという 3 つの観点からディベート学習を支援できることが示された。特に, 従来の振り返りでは議論の内容に関する局所的な分析が中心であったのに対し, 本手法により「誰が誰に反論していたか」「反論のバランスは取れているか」といった構造的な観点からの振り返りが可能となった点は, 本研究の重要な貢献である。

6.2 限界と展望

本研究にはいくつかの限界が存在する。

第一に, 実験参加者の規模と多様性に関する限界がある。DebaTube の評価実験では分析対象が 2 名に限定され, DebaTube Live の実験でも 24 名の参加者による 6 試合という規模であった。より大規模かつ多様な経験レベルの参加者を対象とした利用実験を通じて, システムの効果をより頑健に検証する必要がある。

第二に, 長期利用における効果の持続性と変化に関する限界がある。本研究では単発の練習試合を対象とした実験を行ったが, 継続的な利用がシステムの効果にどのような影響を与えるかは未検証である。長期利用においては, 2 つの相反する可能性が考えられる。一方では, 反論構造の可視化に対する新鮮味が薄れることで, 構造的観点への注意を喚起する力が減退し, 振り返りの質が低下する可能性がある。他方では, システムの操作や可視化の解釈に習熟することで, 初回利用時には気づけなかったより深い構造的特徴の分析が可能となり, 学習効果が向上する可能性もある。今後, 数週間から数ヶ月にわたる長期利用実験を通じて, これらの効果がどのように推移するかを調査し, 継続的な利用においても学習効果を維持・向上させるためのシステムの設計指針を明らかにする必要がある。

第三に, ディベート形式の限定に関する限界がある。本研究では競技ディベートを対象としていたが, この研究が明らかにした反論や構造的議論の特性は, 政治討論や論文の査読などの対立的な議論に限らず, グループディスカッションやビジネスミーティングなどの一般的な議論へも適用可能であると考えられる。今後, 各議論形式の特性や戦略的特徴を考慮した拡張が求められる。

これらの限界を踏まえ, 今後の展望として, 以下のものが挙げられる。

第一に、DebaTube Live の長期利用実験の実施である。本研究では単発の練習試合を対象としたが、数週間から数ヶ月にわたる継続的な利用を通じて、ディベータのスキル向上への長期的効果を検証することが重要である。また、Asian Parliamentary Style への適用も視野に入れ、異なるディベート形式における有効性を比較検討する。

第二に、ジャッジ支援への応用である。本研究ではディベータの学習支援に焦点を当てたが、反論構造の可視化はジャッジの勝敗判定や講評の質向上にも寄与する可能性がある。特に、経験の浅いジャッジが試合全体の議論の流れを把握し、よりの確なフィードバックを提供するための支援ツールとしての展開が期待される。そのため、これまで反論構造の可視化は試合が終わった後に行われていたが、試合中に反論を可視化可能とすることで、どの発言がどの発言に対しての反論になっていたかを確認し、ジャッジを行ううえで重要な議論の流れへの理解が深まる可能性が考えられる。

第三に、リアルタイムな反論構造の可視化がディベータへのフィードバックとして活用可能かの検証である。前述の試合中に行われるリアルタイムな可視化は、特に練習試合においてディベータが議論を組み立てるスキルを向上させるうえで有益な譲歩王となる可能性が考えられる。そのため、ジャッジだけでなくディベータ自身にも試合中に可視化を提示し、学習効果があるか検証することが考えられる。

第四に、他の議論形式への拡張である。本研究では競技ディベートを対象としてきたが、この研究が明らかにした反論や構造的議論の特性は、政治討論やビジネスにおける交渉といった対立的な議論のみならず、グループディスカッションや論文の査読などの協調的な対話へも応用可能であると考えられる。

第五に、実際の競技ディベート環境への普及である。本研究で開発したシステムを、大学のサークルや高校の部活動における体系的なディベート教育の一環として活用することで、より広範な教育的効果の向上が可能となる。そのため第一歩として、DebaTube¹およびDebaTube Live²をインターネット上で公開している。また、今後大規模なアップデートを実施し、世界中のディベータが気軽にディベートの可視化を活用できるようにすることを目指す。

こうした取り組みを通じて、可視化を通じた俯瞰的な議論のフィードバック支援の普及と発展を行い、あらゆるバックグラウンドを持つ人々が建設的な議論を行うスキルを伸ばす手段や機会を提供することを目指す。

¹<https://debatube.nkmr.io/ja>

²<https://debatube.nkmr.io/ja/record>

第7章 結論

本研究では、競技ディベートにおいて俯瞰的な試合特徴の理解が難しいという課題に着目し、反論の構造的な特徴を定量化する手法および、可視化によりディベート学習を支援する手法を提案した。まず、反論構造の定量的な特徴から試合における議論の噛み合い度合いを計算的に評価できるか検証するため、議論の噛み合い度合いに関する仮説に基づく4つの反論構造の特徴量のモデルを定義した。さらに、競技ディベートの試合20試合に対して、手作業による反論判定と熟練したジャッジおよびLLMによる議論の噛み合い度合いの評価を行いコーパスを構築し、これを用いてモデルの精度を検証した。結果として、反論の順序の対応度合いに関するモデルの性能には限界がみられたものの、再反論の頻度、遠い反論の割合、同じ主張への反論の間隔に関するモデルはある程度精度が高いことが分かり、反論構造の定量的な特徴と議論の噛み合い度合いに相関がある可能性が示唆された。

次に、反論構造の視覚的な特徴からディベーターの試合の特徴の理解を支援できるか調査するため、ディベート学習における2つの主要なユースケースを想定し、可視化システムの実装と利用実験によるシステムの効果の検証を行った。まず、参考になる他者の動画視聴を通じて学びを得るユースケースに着目して、反論構造の可視化を動画情報と紐づけて提示することで、複数の試合の特徴の見比べや概観を可能とし、参考になる試合動画の探索を支援するシステム「DebaTube」を実装した。また、本システムを利用して様々な状況で参考になる動画を選択する実験を通じて、反論構造の可視化から議論の特徴を読み取り、効果的かつ効率的な動画探索が促されることを明らかにした。

最後に、練習試合を振り返るユースケースに着目して、試合後に即座に反論構造を可視化し、俯瞰的な振り返りを促すシステム「DebaTube Live」を実装した。そしてシステムの利用実験により、反論の構造的な特徴への着目を促すこと、初心者は議論の流れの整理や想起に反論の可視化を役立てることが多く、上級者は戦略的な判断の正しさの確認や戦略の重要性のチームメイトへの共有に役立てるといったように、様々な層に役立てられる可能性が明らかになった。

以上の結果から、定量化や可視化を通じて反論構造を計算的に扱うことで、ディベーター

タにとって有益な学びを提供できる可能性が明らかになった。競技ディベートを計算的に分析することの利点は主に2つある。1つ目は、ディベート学習機会の普及である。現状、オンラインでの練習会や大会なども行われているものの、ディベート活動が活発な地域とそうでない地域の差は大きく、特に初学者を適切に指導できる先輩や教育者に教えてもらう機会が公平に保証されているとはいいがたい。動画などの豊富な資料は公開されているが、そうした非インタラクティブなコンテンツでは個人の課題への対応力に限界がある。しかし、計算的なディベート学習支援手法は、Webアプリやスマートフォンアプリケーションを通じて、世界中の多くの者が恩恵を受けられ、ディベート学習の質を底上げしえる。

2つ目は、客観性の向上である。先述したようにディベートの資料は豊富に公開されており、また、地域によっては非常に有能な教育者から熱心にディベートの指導を受けることは現状でも可能である。こうした機会は教育的にみても非常に有意義かつ貴重である。しかしながら、そうしたレクチャの大半は定性的なものに著しく偏っており、主観を排除しきれずどうしても偏ってしまう場合や、応用性や一般性に限界がある場合が多くみられる。本研究が切り開いた計算議論学的なディベート学習支援のアプローチが進展し、従来の定性ベースの深いレクチャから得られる学びと組み合わせることで、分析を徹底的に深めつつ、その場その場の話題に左右されすぎずに、議論の流れにそった視野の広いディベートを展開するスキルといった相乗効果の可能性が期待される。

このように、反論構造の定量化と可視化は、計算議論学とディベートの学習支援を融合させ、世界中のディベーターがより俯瞰的に議論を展開するスキルを高めるための基礎を築いたと言える。本研究が、ディベートが論破や不毛な争いを目指すものであるという、とくに日本国内で蔓延している世間の誤解を払拭する一助となり、競技ディベートを教育的にさらに有意義なものとし、ディベーターの人生における意思決定の質を高められるものになることを切に望む。

第8章 謝辞

本論文を執筆するにあたり、ご協力いただいた皆様に心より感謝申し上げます。家族や友人といった周りの方々に温かく見守ってもらえたおかげで、楽しく充実した学生生活を送ることができました。

学部3年から4年間にわたり研究指導してくださった中村聡史先生に深く感謝いたします。先生のご指導のおかげで、国内発表をはじめ、国際会議での登壇発表を経験することができ、やり残したことはないと胸を張って言える研究生生活を送れました。また、根気強く研究の議論に付き合っていたいただいた関西大学の松下光範先生、お会いするたびに新鮮なアイデアを授けてくださったお茶の水女子大学の伊藤貴之先生、研究室の垣根を越えて研究の相談に乗っていただいたり、国際会議の引率を引き受けていただいた小林稔先生、また、実験に協力していただいた Waseda Debate Assosiation のメンバーをはじめとするディベート関係者の方々に、この場をお借りして感謝申し上げます。

また、4年以上、ともに学生生活を歩んでくださった同期の皆様に感謝申し上げます。論文添削をはじめとした研究活動に積極的に協力してくださった先輩方、後輩の皆様にも感謝いたします。日頃から上下関係の壁を越えて関わってくださり、本当にありがとうございました。そして、ここまで応援してくれた家族や周りの方々のおかげで、毎日楽しく過ごすことができました。本当にありがとうございました。

参考文献

- [1] 加藤彰：即興型ディベートの教科書 東大で培った瞬時に考えて伝えるテクニック, あさ出版 (2020).
- [2] Zhang, Y., Yang, X., Feng, S., Wang, D., Zhang, Y. and Song, K.: Can LLMs Beat Humans in Debating? A Dynamic Multi-Agent Framework for Competitive Debate, *arXiv preprint arXiv:2408.04472* (2024).
- [3] Ericson, J., Murphy, J. and Zeuschner, R.: *The Debater's Guide*, Southern Illinois University Press, fourth edition (2011).
- [4] 河野周：中学・高校 英語ディベート入門, 三省堂 (2021).
- [5] Jodoi, K.: The Effects of Parliamentary Debate as a Pedagogy for Argumentation in L1 and L2 Contexts, *Argumentation*, Vol. 39, No. 1, pp. 147–163 (2024).
- [6] Aryanti, M. L., Yuliawati, S., Ekawati, D. and Darmayanti, N.: The Use of Turn Taking in the 2020 US Presidential Debate: A Conversation Analysis Study, *Diglosia: Jurnal Kajian Bahasa, Sastra, dan Pengajarannya*, Vol. 7, No. 3, pp. 511–520 (2024).
- [7] Sari, K. M. and Fatmasari, Y.: Error Analysis on English Diphthongs Pronounced by Indonesian Students in English Debate Competition Posted in YouTube, *Jurnal Dialektika Program Studi Pendidikan Bahasa Inggris*, Vol. 7, No. 1, pp. 38–55 (2019).
- [8] Hadiningrum, I.: Speech Functions Realized by the First Speakers of an English Debate Competition, *J-Lalite: Journal of English Studies*, Vol. 1, No. 1, pp. 53–67 (2020).
- [9] Hartin, P., Birks, M., Bodak, M., Woods, C. and Hitchins, M.: A Debate about the Merits of Debate in Nurse Education, *Nurse Education in Practice*, Vol. 26, pp. 118–120 (2017).
- [10] Mitchell, G. R.: Pedagogical Possibilities for Argumentative Agency in Academic Debate, *Argumentation and Advocacy*, Vol. 35, No. 2, pp. 41–60 (2019).

- [11] Omelicheva, M. Y. and Avdeyeva, O.: Teaching with Lecture or Debate? Testing the Effectiveness of Traditional versus Active Learning Methods of Instruction, *PS: Political Science & Politics*, Vol. 41, No. 3, pp. 603–607 (2008).
- [12] Duffin, F.: Debate Across the Curriculum Results, Technical report, Latitudes in Learning (2005).
- [13] Bedkowski, M. and Rogowska, K.: Persuasive Strategies in Competitive Debates: A Corpus Rhetoric Approach, *Res Rhetorica*, Vol. 11, No. 4, pp. 175–193 (2024).
- [14] Bedkowski, M. and Rogowska, K. J.: Systemic Means of Persuasion and Argument Evaluation: Insights from the Corpus of Competitive Debates, *Informal Logic*, Vol. 44, No. 2, pp. 166–207 (2024).
- [15] National Democratic Institute, : The Youth Leading Debate Instruction Manual (2019).
- [16] Athanasopoulos, C. and Sugino, C. M.: Clash of the Uncivilized: An Alternative Approach to Policy Debate, *Contemporary Argumentation and Debate*, Vol. 38, No. 1, p. 7 (2023).
- [17] Bikakis, A., Flouris, G., Patkos, T. and Plexousakis, D.: Sketching the Vision of the Web of Debates, *Frontiers in Artificial Intelligence*, Vol. 6, p. 1124045 (2023).
- [18] 小林良裕：英語ディベート練習ハンドブック 即興型, S.A.D. Works (2021).
- [19] Ruiz-Dolz, R., Heras, S. and García-Fornes, A.: Automatic Debate Evaluation with Argumentation Semantics and Natural Language Argument Graph Networks, in *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 6030–6040 (2023).
- [20] Hsiao, F.-H., Yen, A.-Z., Huang, H.-H. and Chen, H.-H.: Modeling Inter Round Attack of Online Debaters for Winner Prediction, in *Proceedings of the ACM Web Conference 2022 (WWW '22)*, pp. 2860–2869 (2022).
- [21] Iordanou, K. and Rapanta, C.: "Argue With Me": A Method for Developing Argument Skills, *Frontiers in Psychology*, Vol. 12, p. 631203 (2021).

- [22] Education Endowment Foundation, : Metacognition and Self-Regulation, <https://educationendowmentfoundation.org.uk/education-evidence/teaching-learning-toolkit/metacognition-and-self-regulation> (2023), Accessed: 2025.
- [23] Iordanou, K., Kuhn, D., Matos, F., Shi, Y. and Hemberger, L.: Supporting Strategic and Meta-Strategic Development of Argument Skill: The Role of Reflection, *Metacognition and Learning*, Vol. 17, No. 2, pp. 399–425 (2022).
- [24] Bench-Capon, T. J. M. and Dunne, P. E.: Argumentation in Artificial Intelligence, *Artificial Intelligence*, Vol. 171, No. 10-15, pp. 619–641 (2007).
- [25] Rahwan, I. and Simari, G. R. eds.: *Argumentation in Artificial Intelligence*, Springer, Boston, MA (2009).
- [26] Walton, D.: Argumentation Theory: A Very Short Introduction, in Simari, G. R. and Rahwan, I. eds., *Argumentation in Artificial Intelligence*, pp. 1–22, Springer (2009).
- [27] Ruiz-Dolz, R., Heras, S. and Garcia-Fornes, A.: Structured Argumentation: A Survey, *Artificial Intelligence Review*, Vol. 54, No. 6, pp. 4275–4346 (2021).
- [28] Palau, R. M. and Moens, M.-F.: Argumentation Mining: The Detection, Classification and Structure of Arguments in Text, in *The 12th International Conference on Artificial Intelligence and Law, Proceedings*, pp. 98–107 (2009).
- [29] Peldszus, A. and Stede, M.: From Argument Diagrams to Argumentation Mining in Texts: A Survey, *International Journal of Cognitive Informatics and Natural Intelligence (IJCINI)*, Vol. 7, No. 1, pp. 1–31 (2013).
- [30] Lippi, M. and Torroni, P.: Argumentation Mining: State of the Art and Emerging Trends, *ACM Transactions on Internet Technology*, Vol. 16, No. 2, pp. 10:1–10:25 (2016).
- [31] Stede, M. and Schneider, J.: *Argumentation Mining*, Vol. 11 of *Synthesis Lectures on Human Language Technologies*, Morgan & Claypool Publishers (2018).
- [32] Lawrence, J. and Reed, C.: Argument Mining: A Survey, *Computational Linguistics*, Vol. 45, No. 4, pp. 765–818 (2019).

- [33] Toulmin, S. E.: *The Uses of Argument*, Cambridge University Press, Cambridge (1958).
- [34] Mann, W. C. and Thompson, S. A.: *Rhetorical Structure Theory: Toward a Functional Theory of Text Organization*, Text-Interdisciplinary Journal for the Study of Discourse (1988).
- [35] Moens, M.-F., Boiy, E., Palau, R. M. and Reed, C.: Automatic Detection of Arguments in Legal Texts, in *The Eleventh International Conference on Artificial Intelligence and Law*, pp. 225–230 (2007).
- [36] Stab, C. and Gurevych, I.: Identifying Argumentative Discourse Structures in Persuasive Essays, in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 46–56 (2014).
- [37] Stab, C. and Gurevych, I.: Parsing Argumentation Structures in Persuasive Essays, *Computational Linguistics*, Vol. 43, No. 3, pp. 619–659 (2017).
- [38] Eger, S., Daxenberger, J., Stab, C. and Gurevych, I.: Neural End-to-End Learning for Computational Argumentation Mining, in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 11–22 (2017).
- [39] Ruiz-Dolz, R., Heras, S., Alemany, J. and García-Fornes, A.: Transformer-Based Models for Automatic Identification of Argument Relations: A Cross-Domain Evaluation, *IEEE Intelligent Systems*, Vol. 36, No. 6, pp. 62–70 (2021).
- [40] 栗林樹生, 大内啓樹, 井之上直也, 鈴木潤, 三好利昇, 乾健太郎: 論述構造解析におけるスパン分散表現, *自然言語処理*, Vol. 27, No. 4, pp. 753–779 (2020).
- [41] Opitz, J. and Frank, A.: Dissecting Content and Context in Argumentative Relation Analysis, in *Proceedings of the 6th Workshop on Argument Mining*, pp. 25–34 (2019).
- [42] Zhang, J., Kumar, R., Ravi, S. and Danescu-Niculescu-Mizil, C.: Conversational Flow in Oxford-style Debates, in *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2016)*, pp. 136–141 (2016).

- [43] Egawa, R., Morio, G. and Fujita, K.: Corpus for Modeling User Interactions in Online Persuasive Discussions, in *Proceedings of the Twelfth Language Resources and Evaluation Conference (LREC)*, pp. 1135–1141 (2020).
- [44] Durmus, E. and Cardie, C.: A Corpus for Modeling User and Language Effects in Argumentation on Online Debating, in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 602–607 (2019).
- [45] Ruiz-Dolz, R., Nofre, M., Taulé, M., Heras, S. and García-Fornes, A.: VivesDebate: A New Annotated Multilingual Corpus of Argumentation in a Debate Tournament, *Applied Sciences*, Vol. 11, No. 15, p. 7160 (2021).
- [46] Dung, P. M.: On the Acceptability of Arguments and Its Fundamental Role in Nonmonotonic Reasoning, Logic Programming and n-Person Games, *Artificial Intelligence*, Vol. 77, No. 2, pp. 321–357 (1995).
- [47] Mirkin, S., Jacovi, M., Lavee, M., Kuo, H.-K., Thomas, S., Sager, L., Kotlerman, L., Venezian, E. and Slonim, N.: A Recorded Debating Dataset, in *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC)* (2018).
- [48] Mirkin, S., Moshkovich, G., Orbach, M., Kotlerman, L., Kantor, Y., Lavee, T., Jacovi, M., Bilu, Y., Aharonov, R. and Slonim, N.: Listening Comprehension over Argumentative Content, in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 719–724 (2018).
- [49] Lavee, T., Orbach, M., Kotlerman, L., Kantor, Y., Gretz, S., Dankin, L., Jacovi, M., Bilu, Y., Aharonov, R. and Slonim, N.: Towards Effective Rebuttal: Listening Comprehension Using Corpus-Wide Claim Mining, in *Proceedings of the 6th Workshop on Argument Mining*, pp. 58–66 (2019).
- [50] Orbach, M., Bilu, Y., Gera, A., Kantor, Y., Dankin, L., Lavee, T., Kotlerman, L., Mirkin, S., Jacovi, M., Aharonov, R. and Slonim, N.: A Dataset of General-Purpose Rebuttal, in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pp. 5621–5631 (2019).

- [51] Orbach, M., Bilu, Y., Toledo, A., Lahav, D., Jacovi, M., Aharonov, R. and Slonim, N.: Out of the Echo Chamber: Detecting Countering Debate Speeches, in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 7073–7086 (2020).
- [52] Walton, D., Reed, C. and Macagno, F.: *Argumentation Schemes*, Cambridge University Press (2008).
- [53] Visser, J., Lawrence, J., Wagemans, J. H. and Reed, C.: Revisiting Computational Models of Argument Schemes: Classification, Annotation, Comparison, in *Computational Models of Argument*, pp. 313–324 (2018).
- [54] Wachsmuth, H., Syed, S. and Stein, B.: Retrieval of the Best Counterargument Without Prior Topic Knowledge, in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 241–251 (2018).
- [55] Shi, H., Cao, S. and Nguyen, C.-T.: Revisiting the Role of Similarity and Dissimilarity in Best Counter Argument Retrieval, in *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 12865–12877 (2023).
- [56] Wambsganss, T., Kueng, T., Soellner, M. and Leimeister, J. M.: ArgueTutor: An Adaptive Dialog-Based Learning System for Argumentation Skills, in *Proceedings of the 2021 CHI conference on human factors in computing systems*, pp. 1–13 (2021).
- [57] Guerraoui, C., Reiser, P., Inoue, N., Mim, F. S., Naito, S., Choi, J., Robbani, I., Wang, W. and Inui, K.: Teach Me How to Argue: A Survey on NLP Feedback Systems in Argumentation, in *Proceedings of the 10th Workshop on Argument Mining (ArgMining 2023)*, pp. 19–34 (2023).
- [58] Naito, S., Sawada, S., Nakagawa, C., Inoue, N., Yamaguchi, K., Shimizu, I., Sultana, F., Singh, K. and Inui, K.: TYPIC: A Corpus of Template-Based Diagnostic Comments on Argumentation, in *Proceedings of the 13th Language Resources and Evaluation Conference (LREC)*, pp. 5916–5928 (2022).

- [59] Mim, F. S., Inoue, N., Naito, S., Singh, K. and Inui, K.: LPAttack: A Feasible Annotation Scheme for Capturing Logic Pattern of Attacks in Arguments, in *Proceedings of the 13th Language Resources and Evaluation Conference (LREC)*, pp. 2446–2459 (2022).
- [60] El-Assady, M., Gold, V., Acevedo, C., Collins, C. and Keim, D.: ConToVi: Multi-Party Conversation Exploration Using Topic-Space Views, *Computer Graphics Forum*, Vol. 35, No. 3, pp. 431–440 (2016).
- [61] South, L., Schwab, M., Beauchamp, N., Wang, L., Wihbey, J. and Borkin, M. A.: DebateVis: Visualizing Political Debates for Non-Expert Users, *IEEE Transactions on Visualization and Computer Graphics*, Vol. 27, No. 2, pp. 527–537 (2020).
- [62] Ullmann, T. D., De Liddo, A. and Bachler, M.: A Visualisation Dashboard for Contested Collective Intelligence: Learning Analytics to Improve Sensemaking of Group Discussion, *RIED. Revista Iberoamericana de Educación a Distancia*, Vol. 22, No. 1, pp. 41–80 (2019).
- [63] Carneiro, G., Nacenta, M., Toniolo, A., Mendez, G. and Quigley, A. J.: Deb8: A Tool for Collaborative Analysis of Video, in *Proceedings of the 2019 ACM International Conference on Interactive Experiences for TV and Online Video (TVX '19)*, pp. 47–58 (2019).
- [64] Sevastjanova, R., El-Assady, M., Bradley, A. J., Collins, C., Butt, M. and Keim, D. A.: VisInReport: Complementing Visual Discourse Analytics Through Personalized Insight Reports, *IEEE Transactions on Visualization and Computer Graphics*, Vol. 28, No. 1, pp. 251–261 (2022).
- [65] Gruber, M., Kikteva, Z., Rutter, I. and Hautli-Janisz, A.: DebArgVis: An Interactive Visualisation Tool for Exploring Argumentative Dynamics in Debate, in *Proceedings of the 12th Workshop on Argument Mining*, pp. 140–146 (2025).
- [66] Moon, K., Ha, H. and Lee, K.: Design of Visual Interfaces for Debate Analysis: Co-occurrence Matrix and Circle Packing, *Archives of Design Research*, Vol. 38, No. 1, pp. 167–183 (2025).

- [67] Chen, Q., Wang, Y., Yu, Y., Zhu, X., Yu, X. and Wang, R.: Conch: Competitive Debate Analysis via Visualizing Clash Points and Hierarchical Strategies, *IEEE Transactions on Visualization and Computer Graphics* (2025), Early Access.
- [68] Wambsganss, T., Niklaus, C. M., Cetto, M., Söllner, M., Handschuh, S. and Leimeister, J. M.: AL: An Adaptive Learning Support System for Argumentation Skills, in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–14 (2020).
- [69] Xia, M., Zhu, Q., Wang, X., Nie, F., Qu, H. and Ma, X.: Persua: A Visual Interactive System to Enhance the Persuasiveness of Arguments in Online Discussion, *Proc. ACM Hum.-Comput. Interact.*, Vol. 6, No. 2, pp. 1–30 (2022).
- [70] Slonim, N., Bilu, Y., Alzate, C., Bar-Haim, R., Bogin, B., Bonin, F., Choshen, L., Cohen-Karlik, E., Dankin, L., Edelstein, L., Ein-Dor, L., Friedman-Melamed, R., Gavron, A., Gera, A., Gleize, M., Gretz, S., Gutfreund, D., Halfon, A., Hershovich, D., Hoory, R., Hou, Y., Hummel, S., Jacovi, M., Jochim, C., Kantor, Y., Konopnicki, D., Kons, Z., Kotlerman, L., Krieger, D., Lahav, D., Lavee, T., Levy, R., Liberman, N., Mass, Y., Menczel, A., Mirkin, S., Moshkovich, G., Ofek-Koifman, S., Orbach, M., Rabinovich, E., Rinott, R., Shechtman, S., Sheinwald, D., Shnarch, E., Shnayderman, I., Soffer, A., Spector, A., Sznajder, B., Toledo, A., Toledo-Ronen, O., Venezian, E. and Aharonov, R.: An Autonomous Debating System, *Nature*, Vol. 591, No. 7850, pp. 379–384 (2021).
- [71] Bar-Haim, R., Ein-Dor, L., Orbach, M., Venezian, E. and Slonim, N.: Advances in Debating Technologies: Building AI That Can Debate Humans, in *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing: Tutorial Abstracts*, pp. 1–5 (2021).
- [72] Sviridova, E., Cabrio, E. and Villata, S.: Mining Implicit Arguments for Reasoning: A Survey, *Argument & Computation* (2025), Pre-press.
- [73] Liang, J., Ye, R., Han, M., Lai, R., Zhang, X., Huang, X. and Wei, Z.: Debatix: Multi-Dimensional Debate Judge with Iterative Chronological Analysis Based on

- LLM, in *Findings of the Association for Computational Linguistics: ACL 2024*, pp. 14575–14595 (2024).
- [74] Liu, X., Liu, P. and He, H.: An Empirical Analysis on Large Language Models in Debate Evaluation, in *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics, Volume 2*, pp. 470–487 (2024).
- [75] Schön, D. A.: *The Reflective Practitioner: How Professionals Think in Action*, Basic Books, New York (1983).
- [76] Kolb, D. A.: *Experiential Learning: Experience as the Source of Learning and Development*, Prentice Hall, Englewood Cliffs, NJ (1984).
- [77] García-González, L., Moreno, M. P., Moreno, A., Gil, A. and Del Villar, F.: Effectiveness of a Video-Feedback and Questioning Programme to Develop Cognitive Expertise in Sport, *PLoS ONE*, Vol. 8, No. 12, p. e82270 (2013).
- [78] Groom, R., Cushion, C. J. and Nelson, L. J.: The Impact of Video Feedback on Professional Youth Football Coaches’ Reflection and Practice Behaviour: A Longitudinal Investigation of Behaviour Change, *Reflective Practice*, Vol. 13, No. 4, pp. 475–494 (2012).
- [79] Wang, M., Yuan, B., Kirschner, P. A., Kushniruk, A. W. and Peng, J.: Reflective Learning with Complex Problems in a Visualization-Based Learning Environment with Expert Support, *Computers in Human Behavior*, Vol. 87, pp. 406–418 (2018).
- [80] Schmidgall, S. P., Eitel, A. and Scheiter, K.: Why Do Learners Who Draw Perform Well? Investigating the Role of Visualization, Generation and Externalization in Learner-Generated Drawing, *Learning and Instruction*, Vol. 60, pp. 138–153 (2019).
- [81] Liu, N. F., Lin, K., Hewitt, J., Paranjape, A., Bevilacqua, M., Petroni, F. and Liang, P.: Lost in the Middle: How Language Models Use Long Contexts, *Transactions of the Association for Computational Linguistics*, Vol. 12, pp. 157–173 (2024).
- [82] Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., Hesse, C.,

- Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I. and Amodei, D.: Language Models are Few-Shot Learners, in *Advances in Neural Information Processing Systems*, Vol. 33, pp. 1877–1901 (2020).
- [83] Press, O., Zhang, M., Min, S., Schmidt, L., Smith, N. A. and Lewis, M.: Measuring and Narrowing the Compositionality Gap in Language Models, in *Findings of the Association for Computational Linguistics: EMNLP 2022*, pp. 5687–5711 (2022).
- [84] Bsharat, S. M., Myrzakhan, A. and Shen, Z.: Principled Instructions Are All You Need for Questioning LLaMA-1/2, GPT-3.5/4, *arXiv preprint arXiv:2312.16171* (2024).
- [85] Fukui, M. and Nakamura, S.: Structural Analysis of Rebuttals to Evaluate Argumentative Interaction in Parliamentary Debates, in *Collaboration Technologies and Social Computing*, p. 268–275 (2025).

研究業績

- [1] 福井雅弘, 高久拓海, 中村聡史, 山中祥太. ドライビングシミュレータにおける道路幅の変化が運転に及ぼす影響, 信学技報, Vol.122, No.440, MVE2022-54, pp.33–37 (2023).
- [2] Masahiro Fukui, Takumi Takaku, Satoshi Nakamura, Shota Yamanaka. Evaluating the Applicability of GUI-based Steering Laws to Virtual Reality Car Driving: A Case of Width-Changing Paths, 35th Australian Conference on Human-Computer Interaction (HCI), pp.316–323 (2023).
- [3] 福井雅弘, 中村聡史. 即興型ディベートにおける大局的な反論構造の可視化に基づく議論の噛み合い度合いの基礎検討, 情報処理学会 研究報告コラボレーションとネットワークサービス (CN) , Vol.2024-CN-121, No.31, pp.1–8 (2024).
- [4] 飯田空, 福井雅弘, 高久拓海, 中村聡史, 山中祥太. ドライビングシミュレータにおける路上駐車による道路幅の変化が運転に及ぼす影響, 信学技報, Vol.123, No.433, MVE2023-47, pp.29–34 (2024).
- [5] 成瀬詩織, 飯田空, 福井雅弘, 高久拓海, 中村聡史, 山中祥太. ドライビングシミュレータにおける路上駐車の際のドア開閉が認知的狭さに及ぼす影響について, H 肯定側後半チームシンポジウム 2024, No.B-4-6 (2024).
- [6] 福井雅弘, 中村聡史. DebaTube: 競技ディベートの試合動画探索を支援する大局的な反論構造可視化手法の提案, 情報処理学会 研究報告コラボレーションとネットワークサービス (CN) , Vol.2025-CN-124, No.17, pp.1–8 (2025).
- [7] Masahiro Fukui, Satoshi Nakamura. Structural Analysis of Rebuttals to Evaluate Argumentative Interaction in Parliamentary Debates, The 31st International Conference on Collaboration Technologies and Social Computing (CollabTech2025), Vol.16204, pp.268–275 (2025).