

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

福井 雅弘^{1,a)} 中村 聡史¹

概要: 競技ディベートにおいて、味方に反論が集中した際にどうカバーするか、味方と反論先をどう分担するかといった反論関係の構造への理解を深めるうえで、他者の試合が参考になる。他者の事例を参考にするには試合動画の視聴が有効だが、現状では動画探索に利用できる情報が論題や議論の内容に限られており、試合の反論構造を基に動画を探索する手段が存在しない。その結果、反論の構造を意識して動画を視聴することが難しくなっている。この課題を解決するには、複数の試合における反論構造の特徴を俯瞰的に把握可能とすることが効果的であると考えられる。そこで本研究では、反論構造を可視化するグラフとともに複数の試合動画を提示する動画探索システムを提案する。システムの利用実験を実施した結果、動画探索の目的によって行動の傾向が変わる可能性や、異なる試合でも同じ場面を見比べる傾向がみられることが明らかになった。

キーワード: 競技ディベート, 反論構造, 可視化, 動画探索

1. はじめに

競技ディベートでは、与えられた論題に対して数名の肯定側チームと否定側チームに分かれ、ターン制で議論を行い審判を説得する力を競う。この競技の目的は単に相手を言い負かすことではなく、互いの主張や反論の意図を汲み取り建設的に議論を進めることである。しかし試合形式の性質上、限られた時間で複数の論点を同時に扱わなければならず、議論が複雑に入り組む傾向がある。そのため、建設的な対話に重要な議論全体を俯瞰的に把握することは難しく、軽視されてしまう傾向がある。

このような複雑な議論構造を理解し、適切に対処する方法への理解を深めるには、他者の試合を観察することが有効である。特に参考になる試合動画の視聴は、反論のバランスや一貫性のあるスピーチの構成、チームの役割分担の方法などを学ぶ上で役立つ。しかし、Debate Motions[3]などの既存のディベート動画検索サービスでは試合の論題や大会名などを手がかりに探索することしかできず、試合の内容そのものを基に試合動画を探索することは困難である。そのため、ディベート学習者にとって試合展開に応じ

た議論の組み立て方を学ぶことは難しい。この課題を解決するには、試合展開や反論の流れを俯瞰的に把握できるようなシステムが必要であると考えられる。

そこで本研究では、複数の競技ディベート試合における反論構造をグラフとして可視化することで、反論構造を基にした試合動画の探索および視聴を可能とするディベート動画探索システム DebaTube を提案する。本システムでは、試合の音声の文字起こしから反論構造を可視化し、競技ディベートの試合動画と紐づいた形で提示する。また、本システムの有効性を検証するために、反論の構造を意識した動画探索がどの程度促進されるかについての評価実験を実施する。具体的には、ユーザーが反論構造を理解するうえでユースケースごとにどういった動画探索行動がみられるかを検証し、システムがもたらす効果を明らかにする。

本研究の貢献は以下の通りである。

- 反論構造の可視化を提示することで試合動画探索を支援するシステムの実現
- 競技ディベート経験者を対象としたシステム利用実験の結果から、異なる動画の同じシーンを見比べる傾向があることを明らかにしたこと

¹ 明治大学
4-21-1 Nakano, Nakano-ku, Tokyo, 164-8525 Japan
^{a)} mitit@jf6.so-net.ne.jp

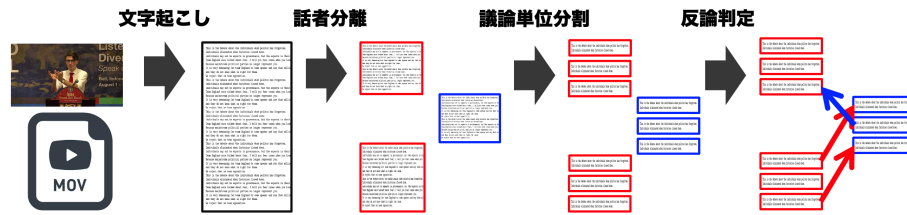


図 1 反論構造の推定手法の概要

2. 関連研究

2.1 議論構造の自動判定手法

ディベートの構造の推定と、それに基づく議論評価タスクを行う研究は幅広く行なわれている。Fa-Hsuan ら [2] は、オンラインの掲示板において、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。これにより、ターンを跨いだ反論関係を学習することで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。また Ramon ら [4] は、準備型と呼ばれるトピックの公開から数ヶ月程度の準備期間を経て試合を行う形式のディベートの試合から議論構造を推定し、議論の質を Argumentation Framework[1] という指標を用いて評価した。結果として、Argumentation Framework と機械学習の両方を組み合わせた手法が高い精度で勝者予測を行える可能性を明らかにした。

また、反論判定の精度向上を目的とした反論評価タスクも行われている。Wachsmuth ら [6] は、オンラインディベート掲示板のコーパスに関して、特定の主張に対する複数の反論候補から、事前にアノテーターの評価により決定した最良の反論を検索する手法を提案した。その結果、平均 7.6 個の候補からトピックに依らず 60%の精度で最良の反論を推定できることを明らかにした。また Shi ら [5] は、同様のタスクを行うための Bipolar-encoder モデルを提案した。実験の結果、モデルが相手側の発言から 74.5%の精度で反論関係を推定できることを明らかにした。

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、他者のディベートが参考になるかという観点で評価タスクを行っている研究はあまりない。また、反論関係を考慮した議論の評価タスクでは扱われる反論の範囲が限定的である。本研究は、試合全体の反論関係の可視化に基づく動画探索を実現することで、より広い視野で試合が参考になるかを判断できるようにすることを旨とするものである。

2.2 可視化による議論構造の理解支援

議論構造の可視化によって論理的議論力を支援する研究はいくつか行なわれている。Wambsganss ら [7] は、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として可視化し、学生が論述文の論理構造を把握する支援を行うシステムを提案した。実験の結果として、システムを利用することで学生の文章の説得力が向上する可能性を明らかにした。また Xia ら [9] は、オンラインディベート掲示板で評価が高い投稿について、論証構造のグラフや論証のタイプの割合を提示することで、他者の論述を参考にするための支援を行なった。実験の結果、システムの利用によりユーザの意見の説得力が向上する可能性が明らかになった。

いずれの研究も議論構造の可視化とその効果の調査に留まっており、可視化の特徴が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しては十分に分析されていない。本研究は、様々なシチュエーションにおいて参考になる競技ディベートの試合動画を探すという実践的なタスクを通じて、ディベートの試合が参考になるかをユーザが判断するプロセスに着目するものである。

3. 競技ディベートの反論構造の自動推定手法

3.1 LLM を用いた反論構造の推定手法

我々はこれまで、2.1 節で述べた議論構造の推定手法に沿って、手作業で反論構造の推定と可視化を行ってきた [8]。しかし、この手法ではディベートの経験者が 1 試合あたり 3 時間程度かける必要があったため、コーパス拡張のスケラビリティに問題があった。また、アノテータの主観に大きく左右されるという問題があった。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 1 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

スピーチに含まれる連続するセグメントをまとめ、Argument Unit (以下 AU と呼ぶ) [4] と呼ばれるひとつの主張や理由づけ、反論などの議論の最小単位とする。2つのスピーチ間の AU のペアを、1つの反論関係として扱う。

セグメント群{segments}を AU 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 AU 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 AU 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 AU に対する反論である AU 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、AU 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 AU のペ

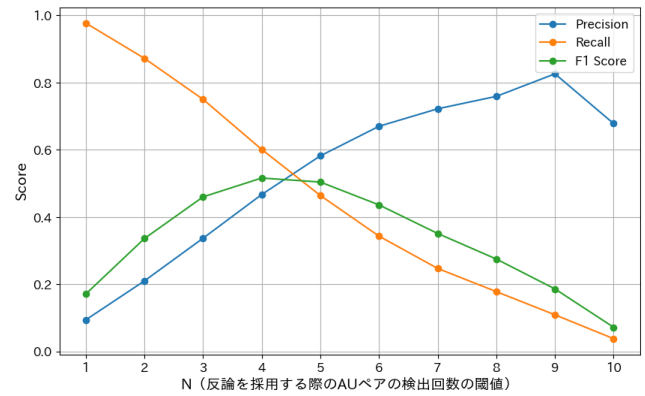


図2 反論を採用する際に用いた AU ペアの検出回数の閾値 (N) と精度の関係

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。ここで、10 回中 N 回以上検出された AU のペアのみを採用する条件で、我々の過去の研究 [8] で手作業で構築したコーパスを正解データとした際の反論元の推定精度を求めた。その結果、を図 2 に示す。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 AU 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. DebaTube

4.1 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と期待できるためである。

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

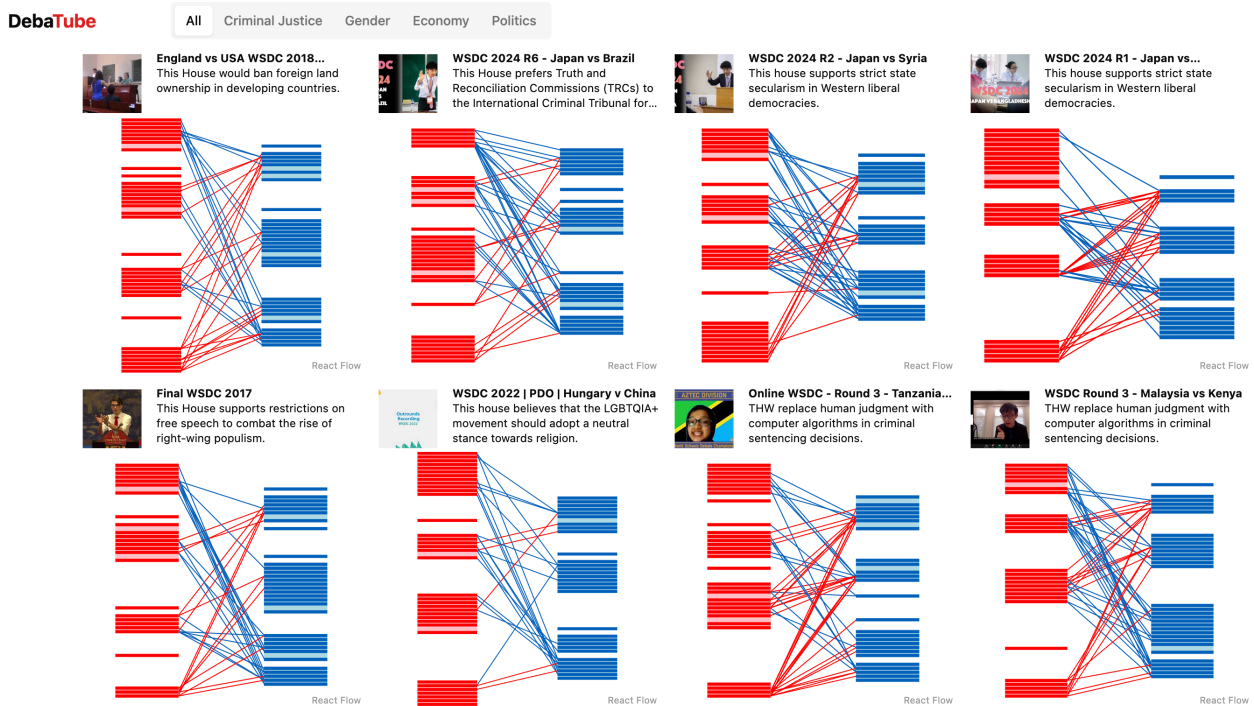


図3 DebaTube システムのユーザインタフェース

められる。

以上の要件をまとめると以下のようになる。

- 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: 各発言と各ターンでのスピーチと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 DebaTube^{*1*2}を実現する。

4.2 実装とシステム概要

図3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。このグラフでは、左側に肯定側の発言ノード、右側に否定側の発言ノードを上から時系列順に配置し、反論関係をノード間のエッジとして表現した。また、どの発言がどちらのチームによって提示されたかを把握しやすくするため、肯定側の発言及び反論を赤、否定側の発言及び反論を青に着色した。ここで、競技

ディベートでは相手側のターン中であっても、ターン開始1分後から終了1分前までの間に15秒ほどの簡潔な質疑を行える「Point of Information」(以下、POIと呼ぶ)というルールがある。これを考慮し、POIが行われている際はノードを相手側に表示し、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。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。

5. 実験

本実験では、「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」を明らかにする。ここでは行動が異なると考えられる、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、可視化がユーザによってどう役立てられるか

^{*1} <https://debatube.nkmr.io>

^{*2} <https://github.com/Non2d/DebaTube>

表 1 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。

を調査する。

5.1 実験の概要

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 1 に、質問の一覧を表 2 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、状況と質問を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。実験はオンライン上で実施し、各参加者が所有している PC で実験を行った。実験では、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 つに取り組むよう依頼した。

5.2 実験用システム

システムは実験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も実装

表 2 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しよう勧めますか？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の一つずつ選び、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しよう勧めますか？特に良いと思われる試合の動画の一つを選び、タイトル左の赤い番号でお答えください。

表 3 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

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

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ったようにした。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ったようにした。

6. 分析と考察

6.1 実験結果

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 3 は参加者ごとのタスクの所要時間、完了タスク数、ノードクリック数である。この表から分かるように、B がスキップしたタスクがある。また、A と C は動画視聴回数が 1 タスクあたり 4 回未満でタスクに十分にに取り組めていない。これはオンラインで実験に取り組んでもらったことが原因で、実験の説明を十分に理解していなかったためと考えられる。そのため、これら合計 3 名を分析対象から除外した。

表 4 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、ピン止め解除数、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平均である。

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。

表 4 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の平均	4.0	4.0	5.5	7.5

表 5 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

6.2 動画探索行動の個人差

前述したように、表 4 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試種類が少ない質問 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確認された。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、つまり最後までピン留めしていた試合数の平均は質問 2 に比べ質問 1 方が少なかったことから、このタスクでは良い動画だけを選ぶタスクと比べて候補が絞られやすいものの、動画を選ぶ理由を固める段階になるまでその傾向がみられない場合があると考えられる。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、参加者によって異なるタイミングで現れる可能性があることがわかった。

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

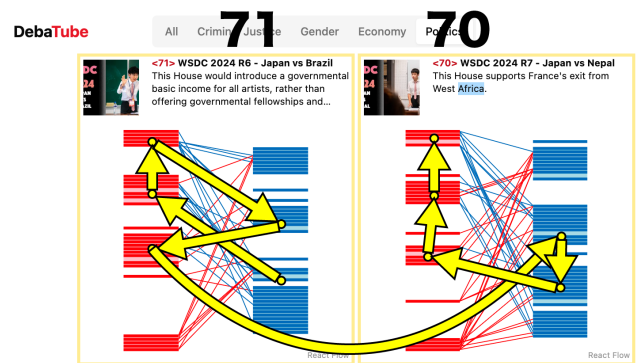


図 4 参加者 D のタスク 4 の開始から 250 秒経過時点でのピン留めされた動画と視聴箇所の軌跡

者間で大きな違いがみられた。表 5 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。

6.3 動画探索行動の詳細な分析

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ターンに着目して動画を視聴する傾向がみられた。図 4 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。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 4 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。最終的に別の動画を選択したものの、異なる動画の同じシーンに着目して見比べる行動がみられた。

同様の傾向は参加者 E でもみられた。図 5 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 人目、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 5 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

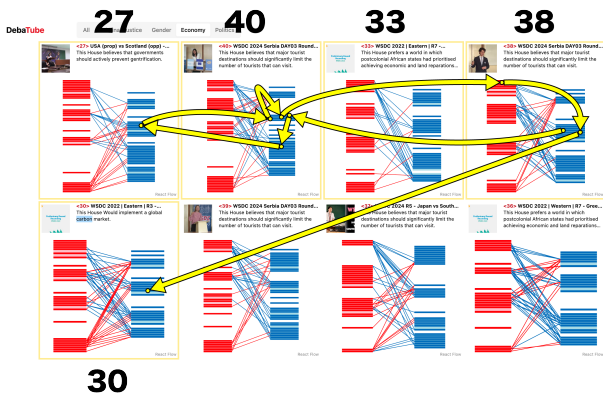


図5 参加者 E のタスク 3 の終了時点での UI とタスク開始 300 秒以降の視聴箇所軌跡

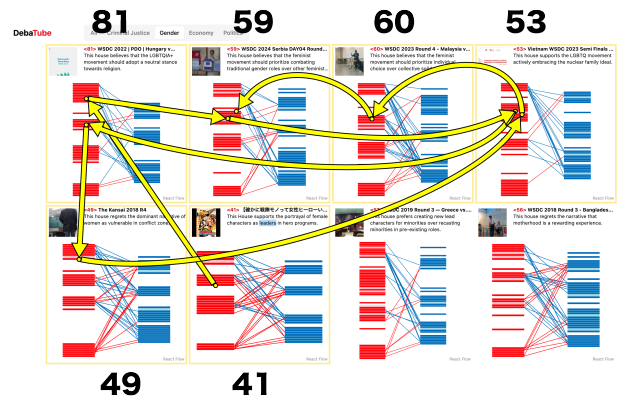


図6 参加者 E のタスク 2 の終了時点での UI と視聴箇所軌跡

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、誤操作であった可能性が高い。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 6)。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の一斉に行った後、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あることが明らかとなった。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着目して探索を行う場合がみられた。

6.4 動画探索の着眼点の個人差

異なる着眼点で動画探索を行っていても、同じ動画を「参考になる良い動画」として選択している事例もみられた。図 7 は参加者 D, E の両者から良い試合として選択さ



図7 両参加者に良い試合と評価された動画

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。

6.5 可視化の役割に関する分析

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相談者に伝えたいメッセージと見つけたい試合の特徴や着眼点を決めて、それに合致する動画を探し、可視化をストーリーリングの材料として用いるパターンである。

本実験では、前者の使用方法が主流であった。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一つのマター（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 は反駁以外の内容が多く、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ど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。

7. おわりに

本研究では、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、ディベート経験者が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開発し、YouTube 上の競技ディベート試合を対象に大規模なデータセットを構築した。その後、競技ディベートの試合動画に対して反論構造を可視化するシステムを実装し、反論関係の特徴を視覚的に提示することで試合内容の理解を支援した。評価実験では、競技ディベート経験者を対象に、特定の条件下での試合動画探索タスクを行ってもらい、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、提案システムを利用することで、反論構造を基に試合動画を効率的かつ効果的に探索できることが確認された。特に、動画探索時のピン留め操作やノードのクリック行動は、反論構造の視覚的理解を通じて試合の特徴を把握する過程を示唆していた。

本研究の成果として、競技ディベートの学習支援システムにおいて、反論構造の視覚化が有効であることが示された。また、本研究の結果は、教育分野における対話型学習支援システムや、議論構造の理解を必要とする他分野の応用にも広がる可能性がある。

今後の展望として、自分の過去の試合データに基づき、

スピーチ傾向の個人差を反映した試合動画探索支援および推薦機能の実現や、British Parliamentary Style と呼ばれる 4 チーム制の形式など、異なるフォーマットの試合への適用可能性の検証が挙げられる。また、反論関係の自動推定精度を向上させるための新しいパイプラインの構築や、リアルタイムな反論構造の推定と可視化機能の実装も行う予定である。

謝辞

本研究の一部は JSPS 科研費 JP22K12338 の助成を受けたものです。

参考文献

- [1] Phan Minh Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77(2):321-357, September 1995.
- [2] Fa-Hsuan Hsiao, An-Zi Yen, Hen-Hsen Huang, and Hsin-Hsi Chen. Modeling inter round attack of online debaters for winner prediction. In *Proceedings of the ACM Web Conference 2022*, volume 45 of WWW '22, page 2860-2869. ACM, April 2022.
- [3] Aaron John. Debate motions. <https://debatedata.io>. アクセス日: 2024-12-11.
- [4] Ramon Ruiz-Dolz, Stella Heras, and Ana García-Fornes. Automatic debate evaluation with argumentation semantics and natural language argument graph networks. *arXiv preprint arXiv:2203.14647*, 2022.
- [5] Hongguang Shi, Shuirong Cao, and Cam-Tu Nguyen. Revisiting the role of similarity and dissimilarity in best counter argument retrieval, 2023.
- [6] Henning Wachsmuth, Shahbaz Syed, and Benno Stein. 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)*. Association for Computational Linguistics, 2018.
- [7] Thiemo Wambsganss, Christina Niklaus, Matthias Cetto, Matthias Söllner, Siegfried Handschuh, and Jan Marco Leimeister. Al: An adaptive learning support system for argumentation skills. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20. ACM, April 2020.
- [8] 中村 聡史 福井 雅弘. 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. In 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), volume 2024-CN-121, pages 1-8, 2024.