

クラウドソーシングによるパラメンタリーディベートへの 論述構造のアノテーション

内藤 昭一^{1,3} 井之上 直也^{2,3} 乾 健太郎^{2,3}

¹ 株式会社リコー ² 東北大学 ³ 理化学研究所 AIP センター

shohichi.naitoh@jp.ricoh.com {naoya-i,inui}@ecei.tohoku.ac.jp

1 はじめに

Argument Mining のタスクの一つとして、論述の構造を推定する論述構造解析がある。論述構造解析は主張や前提などの Argument Component を特定するタスク、Component 間の Support/Attack の関係を分類するタスクなどから構成される [11]。これまでに、Persuasive Essay [10]、Microtexts [8] をはじめとした独話的かつ書き言葉のデータセットを対象に研究が行われてきた。

一方、議論の形態としてディベートや会議のような対話的かつ話し言葉の議論が存在する。対話的な議論は人間のコミュニケーションにおいて中心的な役割を果たしており、これらの議論に対して論述構造解析することは幅広い応用が期待される。そのため、本研究ではそのような議論に対する論述構造解析に着目する。

対話的かつ話し言葉の議論を扱ったコーパスとして、ディベートを対象としたもの [7, 12] や会議を対象としたもの [3, 6] が存在する。しかし、いずれのコーパスにも主張や前提といった Argument Component の論述構造はアノテートされていない。そこで、本研究ではパラメンタリーディベートを題材として、対話的、話し言葉の議論に対する論述構造のアノテートを試みる。

パラメンタリーディベートとは論題に対して肯定、否定のスタンスと話すべき内容の大枠が決められたディベートである。自らの主張と前提、および相手に対する反論が明確に発言されるため、対話的な議論でありながらも一致率の高いアノテーションが期待される。

アノテーションを実施するにあたり、課題となるのは膨大なコストと時間がかかる点である。この課題に対して、クラウドソーシングを利用することで、低コスト、短時間でアノテートをする方法が提案されている [4, 9]。クラウドソーシングで論述構造をアノテートした研究としては、ホテルのレビューコメントを対象としたもの [5] やニュース記事などの WEB 文書を対象としたもの [2] が存在する。これらの研究では、数文からなる文章

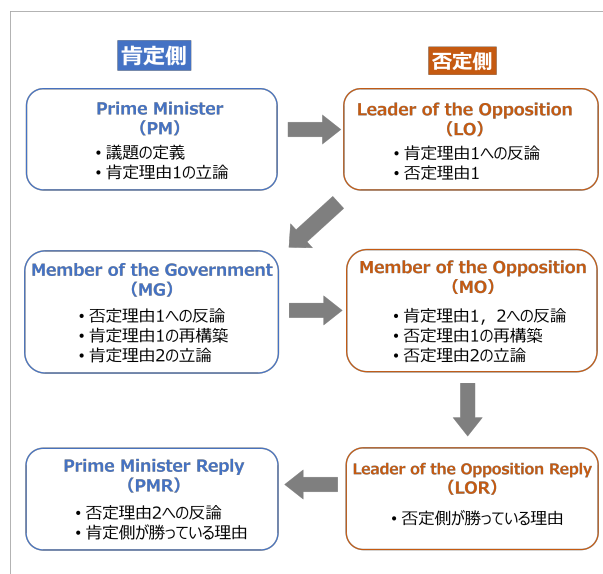


図1: パラメンタリーディベート

をアノテート対象としている。それに対して、ディベートのスピーチは文数が多く、1スピーチ内の Argument Component の数も多いと考えられる。また、他の話者のスピーチに対する反論があるため、1つのディベートの中で内容が連続している。これらの理由から、クラウドソーシングによるアノテーションがより難しいことが予想される。

本研究の貢献は以下の通りである。

- 対話的な議論の論述構造解析に向けて、クラウドソーシングでディベートのスピーチをアノテートした場合の一致度を定量的に検証した
- スピーチのように長く連続した文章に論述構造をアノテートするにあたっての課題を分析し、対策の方向性を示した

2 パラメンタリーディベート

パラメンタリーディベート (以下、PDA) とは、論題が与えられ、肯定側と否定側に分かれて、一般聴衆であるジャッジを説得するディベートである。各チーム3名の計6名で行われ、役割によってスピーチの順番と話

Prime Minister (PM) :

Hi everyone. Today's topic is school lawyers should be sent to all schools. We define the motion as follows. All schools means only elementary school. We have two points. The first point is bullying reduction. The second point is critical [inaudible]. I explain the first point, bullying reduction. We believe that to instill our goals, bullying happens in elementary school because elementary school students are immature. However, they can take responsibility. If something happens, their parent or teacher must take that responsibility. We need to prevent bullying. For that, we need school lawyers. Therefore, school lawyers should be sent to all schools.

Leader of the Opposition (LO) :

Hi everyone. We think that school lawyers shouldn't be sent to all schools. Let me rebut what the government team said. They said that parents or teachers should take responsibility for the problems. However, there still may be many accidents for children because parents or teachers can't go into the problems deeply. So therefore, and more of... Sorry. Let me explain all the points. We have two points. The first point is financial problems. The second point is perspective. How do we explain the first point, financial problems? We believe that much cost must be needed to send school lawyers to all schools.

(以下, 略)

図2: パーラメンタリーディバートのスピーチ例

表1: アノテーションスキーマ

ラベル	説明
論題定義	論題についての定義
主張	論題に対しての賛否
前提	賛否の理由や根拠

すべき内容の大枠が決められている (図1). PDA の実例として, 肯定側 1 人目のスピーカである Prime Minister (以下, PM) と否定側 1 人目のスピーカである Leader of the Opposition (以下, LO) のスピーチを図2に載せた. PM は論題の定義 (波線) を行い, その後に肯定する理由 (下線) と賛成の意 (二重下線) を表明する. そして, 次のスピーカである LO は PM の立論に対して反論を述べた後, 否定する理由を述べる. このように, 役割によってスピーチ内容の大枠が決められた上で議論が進められる.

3 クラウドソーシングによる論述構造のアノテーション

3.1 アノテーションスキーマ

本研究の Argument Component のアノテーションスキーマを表1に示す.*1 今回のタスクでは, ラベル付けする単位は句ではなく文としている. ディバートのスピーチは文数が多いことを考慮し, 文単位のアノテーションをし

*1 本研究の主張は, Stab[10] のアノテーションスキーマにおける Major Claim に相当する. また, 前提は Stab のスキーマにおける Claim と Premise をまとめたものである.

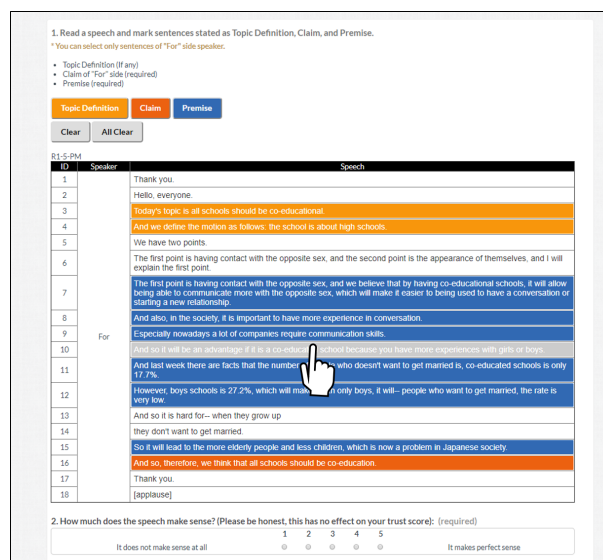


図3: タスクのインターフェース

てから, 後段のタスクで句単位にアノートしていく方針を考えている.

3.2 インターフェース

図3にタスクのインターフェースを示す. ミスの発生を防ぐため, 自由記述方式ではなく, 文を選択するインターフェースとしている. ワークはスピーチを読んだ後, 論題定義, 主張, 前提のいずれかに該当する文を選択し, マークをしていく.

3.3 設定

クラウドソーシングのプラットフォームは Figure Eight (CrowdFlower) を使用する. 1 スピーチあたりの報酬は7セントで, ワークは1回のタスクで5つのス

表2: 論題と件数

論題	件数
体罰は必要か	17
学校弁護士を導入すべきか	25
男女共学にすべきか	25
合計	67

表3: アノテータ間の一致度 (F 値)

論題定義	主張	前提
0.397	0.175	0.307

スピーチにアノテートする。実験では、1 スピーチあたり 10 人のワーカにアノテートしてもらった。

4 実験

4.1 データセット

日本人高校生の PDA スピーチの書き起こしにアノテートをする。なお、スピーチは英語で行われる。ディベートの論題と件数の内訳を表2に示す。

本論文では、スピーチに対してクラウドソーシングで論述構造をアノテートした場合にどのような課題があるのかを探るため、まずはスピーチ内容が比較的単純な Prime Minister のスピーチのみを対象とした。

4.2 結果

図4にアノテート結果の一例を示す。現在のタスク設定がどれだけ難しいものなのか検証するため、アノテータ間の一致度、および Gold との一致度により評価する。

4.2.1 アノテータ間の一致度

アノテータ間の一致度を算出したものを表3に示す。一致度はスピーチ内の論題定義、主張、前提を述べている文を正例、それ以外の文を負例とした場合の F 値により評価する。増ら [1] の手法を参考に、10 人のアノテート結果からペアを取り出し (${}_{10}P_2 = 90$ 個)、ペアの一致度のマイクロ平均を算出している。

表3から、現状では主張に対するアノテータ間の一致度が特に低い結果となっている。主張という概念の捉え方が、論題定義、前提と比べてばらつきがあると言える。

4.2.2 Gold との一致度

ラベルを集約したアノテート結果と Gold との一致度を検証する。Gold は 15 件のスピーチに対して、著者がアノテートしたものとしている。また、集約は、n 人以上が一致した箇所、かつ多数決により最も多かったラベルを採用している。同数だった場合は主張、前提、論題定義の順でラベル付けした。一致人数 n による F 値の

ID	Speaker	Topic	Claim	Premise	Speech
1		0	0	0	Thank you.
2		0	0	0	Hello, everyone.
3		0	0	0	Today's topic is all schools should be co-educational.
4		3	1	1	And we define the motion as follows: the school is about high schools.
5		0	0	1	We have two points.
6		0	1	4	The first point is having contact with the opposite sex, and the second point is the appearance of themselves, and I will explain the first point.
7		1	3	6	The first point is having contact with the opposite sex, and we believe that by having co-educational schools, it will allow being able to communicate more with the opposite sex, which will make it easier to be used to have a conversation or starting a new relationship.
8		0	0	5	And also, in the society, it is important to have more experience in conversation.
9		0	0	4	Especially nowadays a lot of companies require communication skills.
10	For	0	1	5	And so it will be an advantage if it is a co-educational school because you have more experiences with girls or boys.
11		0	1	4	And last week there are facts that the number of people who doesn't want to get married is, co-educational schools is only 17.7%.
12		0	1	4	However, boys schools is 27.2%, which will make-- with only boys, it will-- people who want to get married, the rate is very low.
13		0	0	2	And so it is hard for-- when they grow up
14		0	0	2	they don't want to get married.
15		0	0	7	So it will lead to the more elderly people and less children, which is now a problem in Japanese society.
16		0	6	1	And so, therefore, we think that all schools should be co-education.
17		0	0	0	Thank you.
18		0	0	0	[applause]

図4: アノテーション結果のヒートマップ

変化を図5に示す。主張と前提に関しては、最も良い条件でも F 値 0.6 程度であり、現在のタスクだけで品質の高いコーパスの作成は厳しいと言える。

4.3 課題と今後の対策

Test Question の設定

個別のアノテート結果を分析したところ、適当なアノテートをしているケースが存在していた。品質の低いアノテータを排除する仕組みとして、Test Question がある。Test Question とは答えが分かっている設問を設定することで、その設問に正しく回答できないアノテータを除外する仕組みである。Test Question は完全一致により判定されるため、今回のように選択範囲の完全一致が難しいタスクでは正解となる可能性が低い。

しかし、今回の実験結果から、アノテータ間の一致度とスピーチの文数には相関 (相関係数 $r = -0.40$, $p = 0.80 \times 10^{-3}$) があることが分かった。主張と前提が短く簡潔しているスピーチほど、アノテートの一致度は高い傾向にある。そのため、論題定義、主張、前提が必要最小限で構成されたスピーチに対しては完全一致が期待できる。そのようなスピーチを意図的に作成し、Test Question として設定することで、品質の低いアノテータを除外する効果が期待できる。

タスク分割

同様に個別のアノテート結果から、論題定義、主張、前提を取り違えている理解している事例が存在した。今回のアノテーションは、“文選択”と“選択文へのラベル付与”を同時にアノテートするタスクとなっており、アノテータの認知的負荷が高い可能性が考えられる。Gold との一致度から、主張と前提の判断で特に間違いが発生しやすいと考えられる。タスクを“文選択”と“選択文のラベル付与”に分割し、アノテータの認知的負荷を減ら

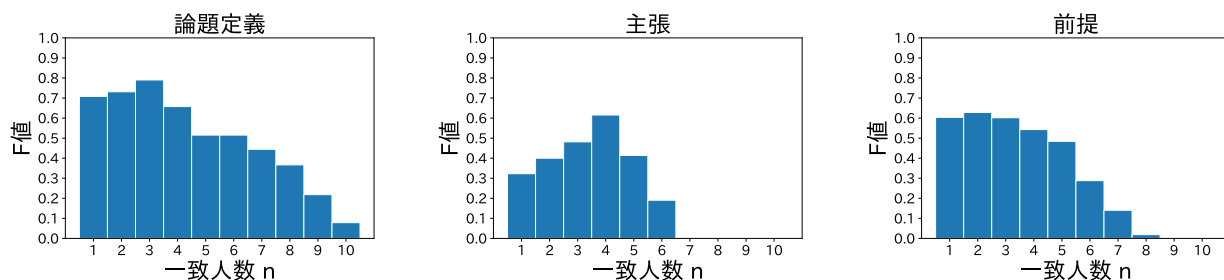


図5: Gold との一致度

すことで、品質を上げられる可能性がある。

5 おわりに

本論文では対話的な議論の論述構造解析に向けて、クラウドソーシングによる論述構造のアノテーションを実施し、結果を定量的に評価した。また、現状のタスク設定の課題を分析し、対策の方向性について示した。

今後は対策の効果の定量的な評価、および Prime Minister 以外のスピーチに対するアノテーションが上げられる。

5.1 謝辞

本研究の一部は JST 未来社会創造事業 (課題番号: JPMJMI17C7) の支援を受けて行った。

参考文献

- [1] Kazuaki Hanawa et al. “A Crowdsourcing Approach for Annotating Causal Relation Instances in Wikipedia”. In: *Proceedings of the 31st Pacific Asia Conference on Language, Information and Computation*. The National University (Phillippines), 2017, pp. 336–345.
- [2] Nguyen Quoc Viet Hung et al. “Argument discovery via crowdsourcing”. In: *VLDB J.* 26.4 (2017), pp. 511–535.
- [3] Adam Janin et al. “The ICSI Meeting Corpus”. In: 2003, pp. 364–367.
- [4] Daisuke Kawahara et al. “Rapid Development of a Corpus with Discourse Annotations using Two-stage Crowdsourcing”. In: *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*. Dublin, Ireland: Dublin City University and Association for Computational Linguistics, 2014, pp. 269–278.
- [5] Mengxue Li et al. “Crowdsourcing Argumentation Structures in Chinese Hotel Reviews”. In: *Proceedings of the 2017 IEEE International Conference on Systems, Man, and Cybernetics (2017)*, pp. 87–92.
- [6] I. McCowan et al. “The AMI meeting corpus”. In: *Proceedings of Measuring Behavior 2005, 5th International Conference on Methods and Techniques in Behavioral Research*. Ed. by L.P.J.J. Noldus et al. Noldus Information Technology, Aug. 2005, pp. 137–140.
- [7] Stefano Menini et al. “Never Retreat, Never Retract: Argumentation Analysis for Political Speeches”. In: *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th Innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018*. 2018, pp. 4889–4896.
- [8] Andreas Peldszus and Manfred Stede. “Joint prediction in MST-style discourse parsing for argumentation mining”. In: *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing* September (2015), pp. 938–948.
- [9] Rion Snow et al. “Cheap and Fast—but is It Good?: Evaluating Non-expert Annotations for Natural Language Tasks”. In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. EMNLP ’08. Honolulu, Hawaii: Association for Computational Linguistics, 2008, pp. 254–263.
- [10] Christian Stab and Iryna Gurevych. “Annotating Argument Components and Relations in Persuasive Essays”. In: *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers (2014)*, pp. 1501–1510.
- [11] Christian Stab and Iryna Gurevych. “Parsing Argumentation Structures in Persuasive Essays”. In: *Computational Linguistics* 43.3 (2017), pp. 619–659.
- [12] Justine Zhang et al. “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*. San Diego, California: Association for Computational Linguistics, 2016, pp. 136–141.