

論述への説明性の高いフィードバック提示に向けたコーパスの試作

内藤昭一^{1,2} 澤田慎太郎³ 中川智皓³ 井之上直也^{2,4} 乾健太郎^{5,2}

¹ 株式会社リコー ² 理化学研究所 ³ 大阪府立大学

⁴ Stony Brook University ⁵ 東北大学

shohichi.naitoh@jp.ricoh.com {szb03072@edu,chihiro@me}.osakafu-u.ac.jp

naoya.inoue.lab@gmail.com inui@ecei.tohoku.ac.jp

1 はじめに

批判的思考力の育成は国家の最重要課題の一つとなっている。従来は人手によるフィードバックに頼っていたが、我々はこれを自動化することを目指している。これにより、人的負荷、生徒の自主学習ができるようになることが期待される。

批判的思考力の育成では、学習者の論述に対して論理構造の観点からのフィードバックを与える。例えば、説得力を高めるために何に言及すれば良いか、立論に噛み合った反論になっているか、論理の飛躍がないか、といったものである。このようなフィードバックを提示する際には、フィードバックだけでなく、その根拠を同時に提示することが効果的だと考えられる。すなわち、システムが学習者の論述構造をどのように解釈し、どの部分が弱いと捉えたのかという判断の過程をもとに提示するということである。例えば、図1のディベートでは、否定側の反論は、肯定側の論理を突くというよりは、単に論題を否定しているだけである。こうした状況に対して、「あなたの反論は肯定側の論理を突いていません」とフィードバックするだけでなく、(1)その論理構造を提示した上で、(2)「家で勉強する習慣のない生徒の成績が下がる」ことが肯定側の立論を攻撃できていないということを指摘し、さらに(3)肯定側の主張「自由に使える時間が増える」に対するデメリットに言及すれば論述を改善できる、ということセットでフィードバックするほうが学習効果が高いと考えられる。

こうしたフィードバックを行うモデルの構築・評価には、論述とその論述構造、及びそれに対するフィードバックを重層的に付与したコーパスが必要である。2節で詳述するが、先行研究では、こうし

論題: 宿題を廃止すべきか?

肯定側の立論

宿題を廃止すべきだ。なぜなら、自由に使える時間が増えるからだ。空いた時間で部活やボランティア、趣味など人生をより充実させる活動ができるようになる。

否定側の反論

宿題を廃止すると、学校で習ったことを復習する機会がなくなる。家で勉強する習慣のない生徒は、成績が下がることにつながる。

否定側の反論 へのフィードバック

立論の「自由時間が増える」という主張に対する直接的な反論がないので、自由時間が増えることによるデメリットに言及すると、より良いでしょう。

フィードバックの根拠

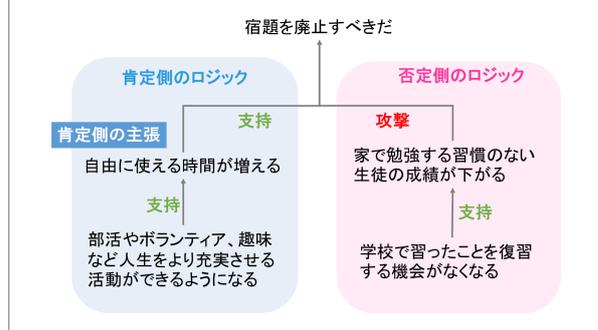


図1 論述へのフィードバックとその根拠の提示

たフィードバックを実現するためのデータセットは存在していない。そこで我々は、一論題について複数の論述を収集し、そこにフィードバックとその根拠となる論述構造を重層的に付与したコーパスを構築することを目指している。

本論文はその最初のステップとして、コーパスの設計方針について論じ、一つの論題に対する試作的アノテーションを行った結果を通して、課題と次のステップについて論じる。また、低コスト・短時間で他の論題へ横展開していくため、クラウドソーシングによる論述構造のアノテーション品質について検証する。

論述構造

主張 (Claim)

However, what students will do during free time is just playing games and chatting using SNS in most cases or just sleeping as they said.

前提 (Premise)

It is because human is lazy in nature. Doing hard thing is usually difficult. Then, many students will not study at home.

支持

フィードバック

改善点 (Suggestion)

ゲームをしたり、チャットをしたりすることがいかに悪いことか、を説明できるとよりよいです

良い点 (Good point)

人間は元来怠けてしまう性質があるという分析はよかったです

図2 論述構造とフィードバックの重層的なアノテーション

2 関連研究

論述の評価に関する一般的な研究として、エッセイスコアリングがある。[1, 2]は、エッセイに対して、Organization や Argument Strength など特定の観点のスコアをアノテートし、スコアを予測するモデルを提案している。[3]は、評価の観点を包括的に調査し、15個の観点についてスコアをアノテートしたコーパスを作成している。[4]は Persuasive Essay Corpus に対して、[5]は ICLE に対して、複数の観点のスコアをアノテートしている。特に、[4]はエッセイ全体ではなく、論述要素という詳細な単位を対象としている。これらの研究は、特定の観点の評価が定量的に分かる点で優れているが、なぜ改善が必要か、どのように改善すれば良いかを伝えるという点では自然文によるフィードバックの方が適していると考える。そのため、本研究ではフィードバックの提示に焦点を当てる。

論述にフィードバックを与える研究として、ユーザが作成した論述に対して反論を与える研究 [6, 7] が存在する。反論を論理の弱い点への指摘と捉えれば、一種のフィードバックと考えることができる。[6]は、論述が含む詭弁の種類と反論をアノテートしたデータセットを作成し、Sequence-to-Sequence モデルによる反論の生成を検証している。[7]は、論述の集合から反論を検索するアプローチをとっている。同じ論点に対して反対のスタンスをとる論述を反論とし、単語の類似度と非類似度から反論としての適切さをスコアリングするモデルを提案している。しかし、これらの研究では、論述構造を捉えた上で反論の生成、検索が実現できているわけではない。

論述構造を活かしたフィードバックを与える研究

としては、主張を裏付ける前提の不足を指摘する研究が存在する [8, 9]。これらの研究では論述構造を解析し、主張を支持する前提の数が不足している場合に、前提の追加を促すフィードバックを行う。本研究では、説得力を高めるためにはどのような内容の前提に言及すべきかという、より内容に踏み込んだフィードバックの実現を目指す。

3 論述文の収集

本研究では、論述構造に基づくフィードバックを提示するため、図2に示すように、ディベートにおける反論に対して論述構造とフィードバックを重層的に付与したコーパスを構築する。本節では、まず付与の対象とする論述文について説明する。4節、5節では、論述構造・フィードバックのアノテーションの仕様および試験的なアノテーション結果についてそれぞれ報告する。

3.1 対象とする論述文

論述文の種類: ディベートにおける反論を付与対象とする。反論を述べるためには、まず相手の論理構造を理解し、これに対してどこに反論するのが効果的かを見極めた上で、さらに説得力のある反論を構築していく能力が要求される。こうした対話的な議論の練習は批判的思考力の育成のための恰好の題材となっており [10]、計算機によるフィードバックができるようになることで、幅広い応用が可能になると考えられる。なお、一般的にディベートは複数のチームがある論題について応酬を繰り返す形であるが、本研究では、図1のように肯定側と否定側の2チームに分かれ、肯定側の立論に対して直接反論している反論を対象とする。また、ディベートは一般的に口頭で行われるが、本研究では初めか

ら書き言葉で書かれた論述を対象とする。音声認識誤りの問題を切り分けて考えるためである。

論題あたりのデータ量: 図 1 のような論述構造に基づくフィードバックの自動化を実現するためには、論述構造を正確に解析する必要がある。このためには、論題に典型的に出現する反論の論述構造をカバーするコーパスをつくることが望ましい。既存の論述構造付きのコーパスでは、より多くの論題をカバーする方針が取られているため、ひとつの論題あたりの論述数が少ない傾向にある。しかしながら、これでは論題に頻出する典型的な論述のパターンを補足するのは難しいと考えられる。

そこで本研究では、ひとつの論題あたり、数百件規模の論述を収集する。一つの論題について考えると、頻出する反論のパターンはある程度絞ることができるはずである。例えば、宿題を廃止すべきという論題においては、図 1 のような議論のほか、肯定側の立論:「強制することで勉強が嫌いになる」-否定側の反論:「マネジメントスキルが身につく」や、肯定側の立論:「教員の負担を減らす」-否定側の反論「生徒がどれほど授業を理解したかわからない」といったパターンが頻出すると考えられる。

もちろん、一つの論題に関して重点的にデータを収集する分、他の論題へ横展開するためのコストは他のコーパスに比べて高い。しかし、実際の利用シーンを考えてみると、いくつかの論題についてコーパスを構築できれば直ちに応用が可能な場面は多いと考えられる。例えば、高等学校の授業での利用を考えると、5つの論題があれば十分に利用価値があるかもしれない。クラスごとに論題が異なっている必要はなく、またこれを毎年変更する必要もないからである。また、フィードバックモデルの利用が広がっていくと、新たな論述文が自動的に集まってくるフィードバックモデルの精度が上がるというエコシステムを形成できる可能性も出てくる。

3.2 データ収集

3.1 節で述べた基本方針に適するものとして、一般社団法人パラメンタリーディベート人育成協会¹⁾ (PDA) が提案する即興型パラメンタリーディベート [10] がある。PDA のパラメンタリーディベートは、肯定側 3 名と否定側 3 名に分かれ、与えられた論題に対して英語で議論をする。各スピーカーごとに話すべき内容の大枠が決まっており、立論や

1) <http://pdpda.org/>

表 1 論述要素のタイプ

ラベル名	定義
立場 (Stance)	論題についての賛否の表明
主張 (Claim)	説得の対象となる意見
前提 (Premise)	主張の説得力を高めるための理由
具体例 (Example)	個人的な体験や現実の物事に即して示された例
その他 (Other)	始まりや締めくくりに用いられる

反論、再構築といったパートが設けられている。

論題 “Homework should be abolished” について 5 つの論点の立論を用意し、PDA に所属するスタッフ、Amazon Mechanical Turk のワーカーより、それぞれの立論に対して 100 件ずつ反論を集めた (詳細は付録 A.2 に記載)。内訳は PDA スタッフが 250 件、Amazon Mechanical Turk のワーカーが 250 件で、合計で 500 件となる。

4 論述構造のアノテーション

4.1 アノテーションスキーム

既存の代表的なスキーム [11, 12, 13] は、節を論述要素の最小単位とし、論述要素のタイプ (主張や前提など) と要素間の関係 (支持や攻撃など) をアノテートしている。しかし、フィードバックを与えるという目的に対して、このような粒度の細かい論述構造が必須かは自明でない。論述構造のアノテーションはコストが高いため、より簡略化したスキームで十分なのであれば、そちらの方が望ましい。そこで、本研究では既存のスキームを簡略化したものをベースとし、実際に付与されたフィードバックを踏まえ、必要に応じて詳細化していく方針をとる。

論述要素の単位は文とし、それぞれの論述要素に対して表 1 に示すタイプを付与する。厳密には前提の中にも、前提に対する前提という階層構造が存在する。しかし、本研究ではそのような場合でも、ひとつの前提としてまとめて捉える。また、反論内部の要素間の関係については支持のみをアノテートする。すなわち、攻撃の関係が現れるのは、立論と反論の間だけとなる。

4.2 アノテーションの方法

別の論題へ横展開をするにあたり、クラウドソーシングを利用することで、低コスト・短時間で拡張できるのが望ましい。そこで、論述構造に関して、クラウドソーシングを利用した場合の論述構造のアノテーション品質を検証する。

表2 フィードバックのカテゴリと実例

観点	説明	例
立論との関連性	立論で提示された論点に対して、直接的な反論ができていますか	立論側の「宿題を廃止することで自由時間が増える」ことに対する直接的な反論も考えましょう
重要性	自身の主張がなぜ重要かをうまく伝えられているか	宿題による勉強が重要であることが前提となっていますが、なぜ勉強が大事であるかまで説明できるとよりよいでしょう
実現可能性	提案した代替案が実現性のあるものになっているか	どのように宿題を早く終わらせて自由時間を作るか、具体的な方策が見えるとよりよいでしょう
具体例	具体例を交えて、聴衆がイメージしやすい説明ができていますか	宿題によってマスターできる事例（授業時間で覚えきれない九九など）を追加できるとよいでしょう

4.1 節で述べた論述構造のスキームに従い、収集した 500 件のデータ (表 3) に対してアノテーションを実施する。アノテーションのインターフェースは付録 A.1 に記載している。クラウドソーシングのプラットフォームには Amazon Mechanical Turk を使用する。アノテーションの報酬は 1 件当たり 25 セントとする。また、アノテーションの信頼度を高めるため、1 件当たり 5 人のワーカを割り当てる。

4.3 評価方法

アノテーション品質の評価は、クラウドソーシングで論述構造のアノテーションを試みた先行研究 [14] に則る形で行う。具体的には、ワーカによるアノテーションを MACE [15] で集約した結果と専門家によるアノテーションの結果の一致率を算出する。ただし、先行研究では論述要素の境界をワーカに選択させるタスク設定のため、境界のオーバーラップを考慮した一致率 Krippendorff's α_U を採用している。本研究では論述要素の境界は文単位で固定のため、Krippendorff's α を採用する。

4.4 結果

ランダムに選択した 30 件の反論について Krippendorff's α を算出したところ、0.40 という結果となった。先行研究 [14] では Krippendorff's α_U が 0.41-0.52 であり、単純な比較はできないが迫る結果となっている。しかし、クラウドソーシングだけで完結できるほどの品質にはなく、その点は今後の課題である。

5 フィードバックのアノテーション

5.1 アノテーションスキーム

論述構造をアノテートした上で、主張と前提の各組み合わせに対して、良い点と改善点を付与する。

その際、どこに対するフィードバックなのかを明示するため、フィードバックを論述要素、あるいは要素間の関係に紐づける (図 2)。

5.2 結果

46 件の反論に対して、PDA に所属するディベートの専門家が与えたフィードバックを分析し、カテゴリに分類したものを表 2 に示す。反論ならではの観点としては、立論との関連性についてのフィードバックが見受けられた。また、具体例を用いた説明がされているかという観点があった。即興型ディベートでは事前準備ができないため、ある事実を裏付ける根拠に言及することが難しい。そのため、すくなくとも個人の体験に基づいた主張なのかという観点が重視されているのだと考えられる。

今後、より多くのフィードバックを分析することで、詳細なカテゴリを定義できる可能性がある。実際のフィードバックを十分にカバーし、かつ詳細なカテゴリ分けができれば、フィードバックを提示する問題をカテゴリの分類とテンプレートをベースとした生成問題として解くという方針も考えられる。

6 おわりに

本論文は、フィードバックとその根拠を同時に提示できるシステムの実現に向け、最初のステップとしてコーパスの設計方針について論じた。また、低コスト・短時間で他の論題へ横展開していくため、クラウドソーシングによる論述構造のアノテーション品質について検証を行った。

謝辞

本研究を進めるにあたり、有益な助言をいただきました理化学研究所の Paul Reisert 氏に感謝申し上げます。

参考文献

- [1] Isaac Persing, Alan Davis, and Vincent Ng. Modeling organization in student essays. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, 2010.
- [2] Isaac Persing and Vincent Ng. Modeling argument strength in student essays. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, 2015.
- [3] Henning Wachsmuth, Nona Naderi, Yufang Hou, Yonatan Bilu, Vinodkumar Prabhakaran, Tim Alberdingk Thijm, Graeme Hirst, and Benno Stein. Computational argumentation quality assessment in natural language. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, 2017.
- [4] Winston Carlile, Nishant Gurrupadi, Zixuan Ke, and Vincent Ng. Give me more feedback: Annotating argument persuasiveness and related attributes in student essays. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2018.
- [5] Zixuan Ke, Hrishikesh Inamdar, Hui Lin, and Vincent Ng. Give me more feedback II: Annotating thesis strength and related attributes in student essays. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 2019.
- [6] Paul Reisert, Benjamin Heinzerling, Naoya Inoue, Shun Kiyono, and Kentaro Inui. Riposte! A Large Corpus of Counter-Arguments. *arXiv e-prints*, p. arXiv:1910.03246, October 2019.
- [7] 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)*, 2018.
- [8] Makiko Ida, Gaku Morio, Kosui Iwasa, Tomoyuki Tatumi, Takaki Yasui, and Katsuhide Fujita. Can you give me a reason?: Argument-inducing online forum by argument mining. WWW '19, p. 3545–3549. Association for Computing Machinery, 2019.
- [9] Joonsuk Park and Claire Cardie. A corpus of eRulemaking user comments for measuring evaluability of arguments. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, 2018.
- [10] 中川智皓, 山内克哉, 新谷篤彦. パーラメンタリーディベート (即興型英語ディベート) における議論の整理と評価の一考察. システム制御情報学会論文誌, Vol. 32, No. 12, pp. 446–454, 2019.
- [11] 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.
- [12] Andreas Peldszus and Manfred Stede. An annotated corpus of argumentative microtexts. In *Argumentation and Reasoned Action: Proceedings of the 1st European Conference on Argumentation, Lisbon 2015 / Vol. 2*, 2016.
- [13] Christian Stab and Iryna Gurevych. Parsing argumentation structures in persuasive essays. *Computational Linguistics*, Vol. 43, No. 3, pp. 619–659, September 2017.
- [14] Tristan Miller, Maria Sukhareva, and Iryna Gurevych. A streamlined method for sourcing discourse-level argumentation annotations from the crowd. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 2019.
- [15] Dirk Hovy, Taylor Berg-Kirkpatrick, Ashish Vaswani, and Eduard Hovy. Learning whom to trust with MACE. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2013.

A 付録

A.1 アノテーションのインターフェース

クラウドソーシングにより、論述構造をアノテーションする際に用いたインターフェースを図3に示す。

Topic: Homework should be abolished.

Original argument

We believe that if homework were to be abolished, we could have more free time. As a result, we could do more of what we really wanted like club activities, hobbies, or playing with friends. In my case, I go to tennis club after class until 5:30 pm and then I go to cram school until 8:00 pm. After this full day, I arrive at my home around 8:40 pm to eat dinner and take a shower. At nearly 10:00 pm I start my homework. I have a lot of homework. As a result, I go to bed late at night at nearly 1:00 am in the morning and I don't have the opportunity to sleep for a long period of time. It is not healthy.

Counter-argument

ID	Speech	Relation ID
0	They said that homework robs us of our free time.	
1	However, Students have holidays and two long vacations in a year.	6
2	They are enough time to do our hobbies or play friends.	6
3	We can go trip in a vacation and play friends in holidays.	
4	Also, We can talk with friends at break time in school.	
5	And besides, We may not have such a lot of homework and cram school everyday.	
6	Students have a lot of free time than you think.	
7	At least, they have more free time than adults.	

Stance : statements that express agreement or disagreement with a topic

Claim : central reason given in support of the debater's stance (In general, there is one claim, at most two claims per speech)

Reason : additional reasons that underpin the plausibility of the claim

Example : personal experiences, hypothetical situations, statistics, and references in newspapers, etc.

Other : greetings at the beginning and end, summary at the end such as "That's the reason why ..."

図3 アノテーションのUI

A.2 収集データの詳細

収集したデータに関して、立論の論点と件数の内訳を表3に示す。

表3 論題“Homework should be abolished”について収集した反論の件数

立論の論点	反論の件数
自由時間の増加	100 件
生徒の主体性の向上	100 件
勉強に対する嫌悪感の低減	100 件
誤った勉強法の定着阻止	100 件
家庭にかかる負担の減少	100 件