

オンライン説得議論フォーラムにおける 高粒度なアノテーションスキームの提案

江川 綾^{1,a)} 森尾 学² 藤田 桂英²

概要：近年，Web 議論掲示板には様々な形式のものが登場している．中には他人の意見を変えるチャレンジを行う掲示板，説得議論フォーラムも登場している．現在，この説得議論フォーラムにおいて，説得に効果的な議論の特徴の解析を行う研究がさかんに行われているが，説得議論の特徴を捉え，投稿間のインタラクションを考慮した議論データセットは未だに存在していない．そこで本論文では，説得議論の特徴である，投稿文章の論理関係や修辞技法を捉え，投稿間のインタラクションを考慮したアノテーションスキームを提案する．さらに，提案スキームに基づきコーパスを作成し，作成したコーパスにおいて説得に効果的な議論特徴の解析を行った．結果として，説得性の高さには「事実と意見をどれだけ述べているのか」「修辞技法をどれだけ使っているのか」などは大きな影響はなく，どのように文章を組み立てて説得を行っているのかが重要であり，「事実を根拠として明確に示した上で，行動方針や指針を提示すること」が説得力を向上させる可能性があることが明らかになった．

1. はじめに

近年 Web 上でコミュニケーションを行う機会が増えており，掲示板などオンライン上で人々と議論を交わすことが一般的になってきている．しかしオンラインでの議論には，参加者の能力にばらつきがあることや，背景知識を共有していないこと，参加者が根拠なしに意見を述べることで発散的な議論が発生するなどの問題がある．そこで，こうした問題を解決し議論の質を高めるために，議論の質を定義し評価する研究が近年数多く行われている ([1], [8], [12], [13])．こうした議論の質は主に前提の合理性，主張に至る論理関係の妥当性，信頼性や有効性などによって評価される．前提の合理性や論理関係の妥当性の評価は，文章における議論構造の抽出が重要となるため議論マイニング (Argument Mining) [5] の研究領域と親和性があるといえ，この技術を適用し説得力や主張の強固さ，有効性などを向上させる取り組みがある．

また，オンライン議論掲示板には様々な形式のものがある．例えば，市民が政策に対する意見を議論し合う市民参加型の掲示板や，スレッド主の意見や主張に対して説得

を試みる説得掲示板などがある．こうした掲示板を議論コーパスとして用いて解析を行う研究が多数行われており，特に説得議論掲示板における解析は，相手の意見や価値観に変化を与えることが社会的な活動において重要であるため，有用であるとされている．既存研究において説得議論掲示板の解析を行った例として，説得議論掲示板の ChangeMyView^{*1} を用いて説得に効果的な議論特徴の解析を行ったものがある ([13], [16])．これらの研究では投稿における議論構造ではなく，文章的特徴 (単語や文章スタイル) に着目し特徴の解析を行っている．一方 ChangeMyView において，議論マイニングの研究を適用し，議論構造を踏まえた解析を行う研究も行われている．これらの研究は投稿文章とその順接的な推論関係を捉えたアノテーションを実施し，解析を行っている．しかし説得議論において，粒度が高く，投稿間のインタラクションを捉えたデータセットは依然として存在していない．このことから，説得に効果的な特徴の解析のために，文章の意味的特性，投稿間のインタラクションを捉えた高粒度なアノテーションスキームとそれに基づくコーパスデータの作成が求められている．

本研究では，説得議論掲示板 ChangeMyView を用いて投稿文章の性質を考慮し，

- 文章の意味的特性
- 投稿内の推論関係とその性質

¹ 東京農工大学工学部
Faculty of Engineering, Tokyo University of Agriculture and Technology

² 東京農工大学大学院工学研究院
Institute of Engineering, Tokyo University of Agriculture and Technology

a) s155745v@st.go.tuat.ac.jp

^{*1} <https://www.reddit.com/r/changemyview/>

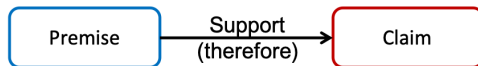


図 1 基本 Argument 構造

- 投稿間のインタラクション

これらを捉えた高粒度なアノテーションスキームの提案を行う。さらに、スキームに基づき作成した説得議論コーパスを解析し、説得における効果的な議論の特徴を明らかにする。

論文の構成を以下に示す。第 2 章では、議論マイニングと議論の質の評価に関する関連研究について述べる。第 3 章では、提案するアノテーションスキームについて述べ、第 4 章でアノテーション結果について述べる。また、第 5 章ではコーパスを解析し、議論要素の説得における役割を明らかにし、最後に 6 章で本稿のまとめを述べる。

2. 関連研究

2.1 議論マイニング (Argument Mining)

自然言語処理の研究領域のひとつとして、議論マイニング (Argument) [5] がある。議論マイニングは哲学、論理、自然言語処理の複合領域であり、議論的な文書の中から主張や前提となっている部分を抽出し、構造化する学問領域である。特に、書き言葉による論理や (logic) や修辞学 (rhetoric) などと密接に結びついており、古代ギリシャの三段論法とも関係が深い。また、議論マイニングは Argument を抽出することに焦点を当てており、将来的にこの技術は文書要約、検索、対話ボット、法的文書解析など様々な領域で応用されると考えられている。

主張のことを英語で Claim、前提のことを Premise という。また主張と前提の構造を英語で Argument という。Palau, Stab, Peldszus らは Argument を「一つの主張と複数の前提 (もしくは前提なし) のペア」と定義した ([6], [9], [10])。前提はその主張に対して正当化を行う要素のことであり、前提は主張の理由となることが多くあるため「前提 (Premise) → therefore → 主張 (Claim)」という関係が成り立つ。最も基本的な Argument 構造は図 1 のようになる。

また、こうした議論構造を捉える議論マイニングのシステムは、様々な Argument 構造のモデルに基づいて構築される。最も有名な Argument 構造のモデルとして Toulmin モデルがある [14]。Toulmin モデルは主張 (Claim)、事実 (Datum, Data)、理由付け (Warrant)、裏付け (Backing)、限定 (Qualifier)、反証 (Rebuttal) からなっている。議論における主な要素としては主張、事実、理由付けの三つであり、これらの要素により議論構造を解釈するモデルである。しかし、このモデルにおいては事実と理由付けが解釈として混同してしまうケースなどがあり、ドメインに応じ

て適切にモデルを考え適用する必要があるとされている。

2.2 Argument Quality Assessment

近年、説得力や議論構造の強さ、受容性など様々な質の尺度に基づきデータセットを構築し、議論マイニングの技術を活用する研究が行われている。

2.2.1 Evaluability

Park らは、議論掲示板においてユーザがデータや理由付け無しで主張を行うケースが多く見られることから、ユーザコメントの評価性を捉える Argument モデルの提案を行った [7]。Park らのモデルは、議論要素の種類を *Fact, Testimony, Value, Policy, Reference* の 5 種類、要素間の関係性を *Reason, Evidence* の 2 種類としてそれぞれ定義し、市民議論掲示板 eRulemaking*2 のデータに対してアノテーションを行った。これは投稿するにあたり議論構造が十分に構築されているかを評価できるモデルとなっており、ユーザコメントの自動評価・フィードバックを行うシステムなどに活用できる。

2.2.2 Persuasiveness

説得議論掲示板 ChangeMyView において、説得に効果的な議論特徴の解析が盛んに行われている ([2], [13], [16])。Hidey らは、アリストテレスが提唱した 3 つの説得手段 (ethos, logos, pathos) を踏まえ、前提と主張を意味的に分類し、その Argument モデルを用いて解析を行った [2]。このモデルは、前提 (Premise) を *ethos, logos, pathos* の 3 種類に、主張 (Claim) を *interpretation, evaluation(rational), evaluation(emotional), agreement, disagreement* の 5 種類に分類しアノテーションを行った。このような粒度の高いアノテーションにより、説得議論を意味的に解析できるとされている。

また、意味的に高粒度な解析を行うことに加え、投稿文章間でどのようなインタラクション (支持/反論等) が生じているのかを明らかにすることが求められている。そこで本研究では説得議論における議論要素の意味的分類に加え、要素間の論理関係、さらには投稿間のインタラクションを捉えたスキームを提案しアノテーションを行う。

3. アノテーションスキーム

3.1 データソース

アノテーションには、Reddit の特定トピックを扱う説得議論フォーラムである ChangeMyView を用いるものとする。ChangeMyView では、参加ユーザの意見や観点 (View) を変えるために説得議論を行う。

- スレッド主が View を投稿する。スレッド主の投稿は Original Post (以下, OP) と呼ばれる。OP では View の内容に加え、その View の背景にある理由や根拠を

*2 <http://www.regulationroom.org>

十分に述べなければならない

- 投稿には十分な情報量を保証するために、500文字以上記述しなければならない
 - 説得成功した場合、スレッド主からチャレンジャーに対して報酬として Delta (Δ) を与えなければならない
- といったルールが存在しており、単なる論争の場ではなく、参加者が様々な意見や価値観を理解することに役立つことを目的としている掲示板であるといえる。そのため、文書において主張 (Claim) と前提 (Premise) の構造化を試みる議論マイニングは有用な解析として期待できる。そこで、ChangeMyView に対して独自の議論マイニングのスキームを適用し、アノテーションを行う。

また、その投稿主の意見や観点 (View) を変えることに成功した投稿 (= Δ を与えられた投稿) のことを *Positive*、変えることが出来なかった投稿のことを *Negative* と呼ぶ。アノテーションにおいては、ChangeMyView を無作為に抽出し作成した Tan らのデータセット [13] を用いるものとし、(*OP*, *Positive*, *Negative*) の3つの投稿を1セットとして処理を行う。

3.2 議論要素の種類

説得議論の意味的特徴を捉えた議論要素を定義した。この定義はユーザコメントの評価性を捉えた Park らの Argument モデル [7] を拡張したものとなっている。5種類の議論要素の定義は以下の通りである。

Fact

「事実に関する表現や記述、または、個人的な感情・偏見・解釈によって歪められることなく知覚される状態に関する客観的な記述」の命題のこと。定義により、Fact は客観的な証拠を以って真偽を検証できる記述のことを指し、真偽を問う必要はない。また、客観的な証拠に基づき検証される必要があるため、将来についての予測は検証不可能であり、Fact であるとはいえない。

例は以下の通りである。

- Empire Theatres in Canada has a “Reel Babies” showing for certain movies
- they did exactly this in the U.K. about thirty or so years ago
- this study(<http://www.businessnewsdaily.com/3150-women-speak-less.html>) shows that women are 75% less likely to speak up in a space when outnumbered by men

Testimony

「著者の個人的な状態や経験に関する記述」の命題のこと。この命題は、個人的な状態や経験に関する事柄であるため、これを証明できる客観的な証拠が存在しないことが

多い。つまり、Testimony においては客観的な証拠が必要ではなく、過去の経験や状態、現在の状態に関する記述が Testimony となる。

例は以下の通りである。

- I've heard suggestions of an exorbitant tax on ammunition.
- I do not have children.
- I live in a city now, and I cannot even take public transportation

Value

「何をすべきかについての具体的な主張ではなく、価値判断を含んだ記述」の命題のこと。価値判断は、意見や偏見、感情などに代表される主観的なものであるため、客観的な証拠で検証することはできない。Value は事実ではなく意見としての性質を持つ記述であるため、一般論らしい論調で記述されている場合でも、客観的に検証することができず、解釈が多様になることが予想される文章は Value となる。

例は以下の通りである。

- this is completely unworkable
- it is absolutely terrifying
- women are capable at doing whatever they want

Policy

「特定の行動方針や指針を提示し、それを促す記述」の命題のこと。多くの場合、文章に *Should* や *ought to* などの助動詞を含む。Policy も Value と同様に、価値判断を含む記述であるため、客観的な証拠で検証することはできない。Value との違いは提示した指針に対して行動を促しているか否かである。

例は以下の通りである。

- we should not permanently punish people for mistakes they made in the past
- even if you don't believe in logic, you had better act as if you do
- intelligent students should be able to see that

Rhetorical Statement

「直接的に意見を述べずに、言い回しを工夫することで相手の感情に訴え、筆者の意見に結びつけるような記述」のこと。Rhetorical Statement においても、Value や Policy と同様に価値判断をもとにした行動指針の提示や意見の表明がなされるが、比喩表現や疑問を投げかけるなどして相手の感情に訴えかけるという点で異なっている。また、価値判断を含む記述であるため、客観的な証拠で検証することはできない。

例は以下の通りである。

- What was the critical difference?
- You can observe this phenomenon yourself!
- if one is paying equal fees to all other students why is one not allowed equal access and how is this a good thing?

3.3 要素間関係の種類

要素間関係の定義はスキームによって様々であり, Stabらは *Support*, *Attack* の2種類を, Kirschnerらは *Support*, *Attack*, *Detail*, *Sequence* の4種類 [3], [11] を定義している. これらをもとに, 提案スキームでは2種類の要素間関係を定義した. 定義は以下の通りである.

Support

「ある要素 X が他の要素 Y に対して, 賛成のスタンスをとる」ような関係を指し, ある要素 X から他の要素 Y が順接的な推論関係によって導かれるとき, X は Y に対して **Support** 関係があるとする. 順接的な推論関係とは, *therefore*, *so* といった接続詞によって接続されるような関係をいう. そのため, Support 関係は前提の理由付けや例示により主張を支援するような関係性となる.

例は以下の通りである.

- **X**: Getting a license means demonstrating some kind of proficiency related to safely and responsibly strong and using the weapon (Value)
Y: If you aren't properly licensed, you can't own or possess a gun (Policy)
- **X**: Downs syndrome is a mutation that causes sterility (Fact)
Y: people with downs syndrome don't have children, and there is no genetic cause for it (Value)

Attack

「ある要素 X が他の要素 Y に対して, 反対のスタンスをとる」ような関係を指し, ある要素 X から他の要素 Y が逆接的な推論関係によって導かれるとき, X は Y に対して **Attack** 関係があるとする. 逆接的な推論関係とは, *however*, *but* といった接続詞によって接続される関係をいう. そのため, Attack 関係は前提が主張に対する反例や反論となるような関係性となる.

例は以下の通りである.

- **X**: Young men are the most likely demographic to get into an accident (Value)
Y: that does not warrant discriminating against every individual in the group (Value)
- **X**: They supported me for a long time (Testimony)
Y: when I look at them objectively, they are average human beings (Value)

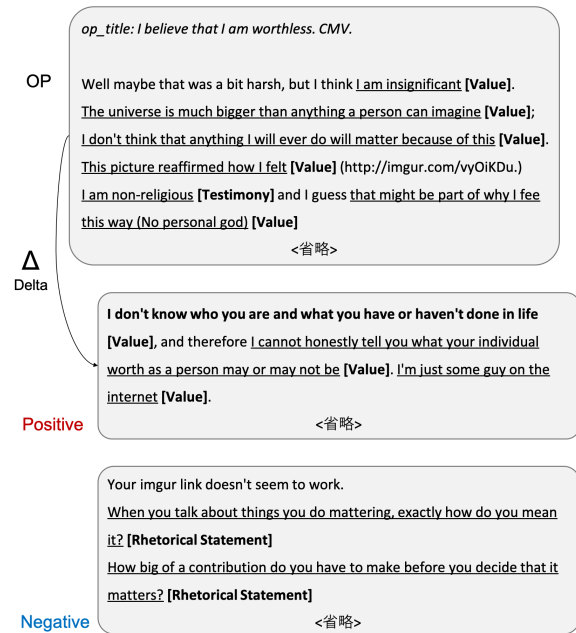


図 2 アノテーション例

4. アノテーション結果

データセットからランダムに 50 セット (= 150 投稿) を抽出し, *OP*, *Positive*, *Negative* それぞれにおいて, 議論要素の種類のアノテーションを実施した. アノテーション実施例を図 2 に示す.

議論要素の種類のアノテーションにおいては, 文章レベルでなく単語レベルで, 定義した 5 種類の議論要素をアノテーションする. 単語レベルのアノテーションでは, 一つの文章中に複数の議論要素が含まれている場合に対応できるため, 最も粒度の高いアノテーションとなる.

アノテーションにより作成したコーパスには, 合計 3759 個の議論要素が含まれている. *OP*, *Positive*, *Negative* それぞれにおける議論要素の種類の内訳は Table1 に示す. Table1 によると, 要素数の合計に対して *Value* が約 65% を, *Rhetorical Statement* が約 16% を占めており, 大半の文章では意見を直接的, もしくは修辞技法を用いながら伝えることで説得を試みていることがわかる.

作成したコーパスにおいては, 各セットに対して 3 人のアノテータ全員が独立してアノテーションを行い, 最終的に 3 人の結果の多数決で議論要素の種類とその範囲を決定する. アノテータ間一緻度は *Krippendorff's α* [4] を用いて計測する. *Krippendorff's α* は, トークンレベルのアノテーションに対応しており, 作成したコーパスに対して有効である. 50 スレッド全体のアノテータ間一緻度を算出した結果, $\alpha = 0.691$ であった. 議論掲示板においてアノテーションを行った Park ら [8] ($\alpha = 0.648$) や, Hidey ら [2] ($\alpha = 0.65$) と比べても高い一緻度であることから有効な

表 1 コーパスにおける議論要素の種類と数

	Value	Policy	Fact	Testimony	Rhetorical Statement	Total
OP	665	38	49	104	127	983
Positive	1075	50	117	125	279	1646
Negative	730	35	72	79	214	1130
All	2470	123	238	308	620	3759

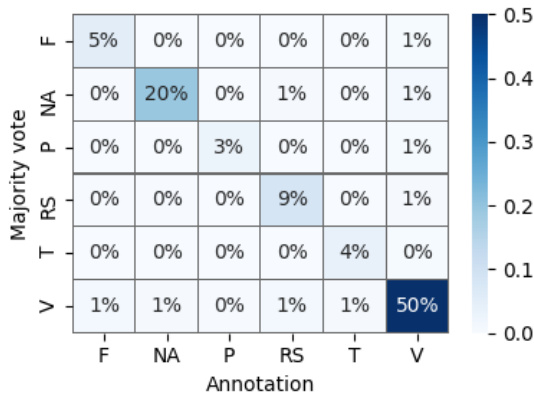


図 3 議論要素の種類のアノテーションにおける混同行列

コーパスである。

また、アノテータ間で一致度が低くなる議論要素の傾向を調べるために、多クラスの混同行列を作成した。混同行列は図 3 に示す。図において、Majority vote が統合後のデータ、Annotation が各アノテータによるアノテーションデータのことを示しており、RS, NA はそれぞれ Rhetorical Statement, 要素なし (No Annotation) を表す。混同行列によると、

- Value と Fact
- Value と Testimony
- Value と Rhetorical Statement

においてアノテータ間の不一致が発生しており、Value と他の種類とで不一致が発生するケースが多い。Value と Fact においては、Value が意見性、Fact が事実性の命題であることが特徴であるが、「多くの人々が」「日本では」といった意見を一般的論に見せかけた記述の際に不一致が発生することが多くなる。また、可能性についての記述に関しても事実か意見で解釈が割れ、不一致となることが多い。Value と Testimony においては、いずれも客観的な証拠により検証可能ではないため、経験と価値判断のいずれとも取れるような記述の際に不一致になる。例えば、「私は無神論者です」という記述は個人の状態の記述であるため Testimony であるが、価値判断・思考の記述とみなし Value とアノテーションするケースが見られた。Value と Rhetorical Statement においては、意見が直接的に述べられているか否かという部分で解釈が分かれば、不一致が発生することがあった。

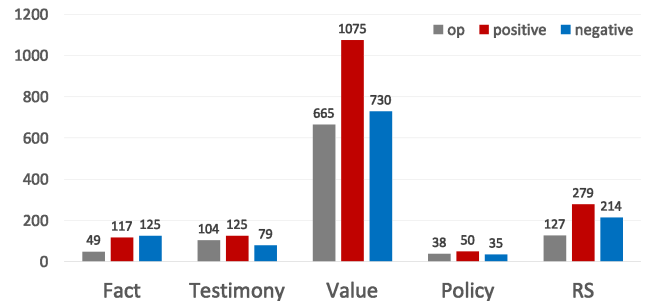


図 4 各投稿における議論要素の分布

5. 説得における議論要素の解析

作成したコーパスを用いて、各議論要素がそれぞれ説得においてどのような役割を持つのか解析を行う。投稿においては OP と Positive/Negative を比較・解析することによって説得を試みる文書の性質、Positive と Negative を比較・解析することで説得議論に効果的な文書の特徴をそれぞれ明らかにする。

5.1 議論要素の数

各投稿における各議論要素の数の分布を図 4 に示す。この議論要素の数について、OP v.s. Positive/Negative, Positive v.s. Negative のそれぞれにおいて検証を行った。結果として、説得を試みる文書は、意見を表明している OP の文書に比べて Testimony が少なく、Rhetorical Statement が多いということが明らかになった。つまり、

- 意見とその理由を表明する際には、個人的な経験などによる理由付けを行うことが多い
- 意見を表明する文に比べ、説得文章のほうが修辞技法をより多く用いる

ということがわかる。

また、Positive と Negative の比較をしたところ、議論要素の数に有意差はない。このことから、議論要素の種類とその数自体は説得に大きな影響がないことがわかり、「事実と意見をどれだけ述べているのか」「修辞技法をどれだけ使っているのか」といった特徴でない要素によって説得性が左右されていることが明らかになった。

5.2 議論要素の位置

各議論要素における、投稿ごとの要素出現位置の分布を図 5 に示す。なお、縦軸の Position は特定要素が文中の要

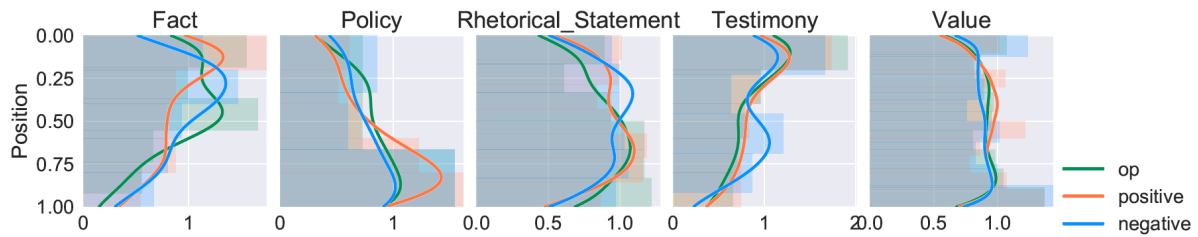


図 5 各投稿における議論要素の出現位置の分布

素数の合計に対してどの位置に出現したのかを表すものとしており、(特定要素の出現位置) / (要素数の合計) で表されるため 0~1 の範囲を取る。そのため図中の縦軸の 0 は投稿の最初の要素, 1 は投稿内の最後の要素であることを表している。

Fact

Positive, Negative の双方において前半部に出現することが多い。特に Positive は他の投稿と比較して、初めに Fact を述べるケースが多く、事実性の客観的な命題をベースに説得を行っている。また、OP, Positive, Negative 共に後半部にて出現するケースは少なく、文書の冒頭で前提として事実を述べるケースが多い。

Testimony

OP, Positive, Negative でほとんど差はない。共通して、文章の冒頭にて自分の経験等の記述を行っており、これは意見の表明や説得を行う上での前提の共有をしているのだと考えられる。Negative は他の投稿と比べて中盤でも Testimony が出現することがあり、これは主張を経験談で補強するケースなどが考えられるが、こうした説得技法は効果的でない。

Value

OP, Positive, Negative でほとんど差はなく、似た形の分布になっている。そのため、意見をどの程度表明しているか自体は投稿の特徴、説得性に対して影響がなく、意見の内容によって左右されていると考えられる。

Policy

OP, Negative でほとんど差はないが、Positive のみ文章の後半に多く出現している。このことから、文章の後半において「特定の行動方針や指針を提示し、それを促す」ことは説得において効果的であると考えられ、最後の後押しとして「~のように考えるべきだ」「~するべきだ」という主張を用いるのは効果的である。

Rhetorical Statement

OP, Positive, Negative で大きな差はない。共通して、文章の中盤から後半にかけて分布が集中しており、前提を述べ、意見を表明した上での後押しとして、相手の感情に訴えかける記述をしているのだと考えられる。

各要素の比較をまとめると以下の通りになる。

- Fact, Testimony は、OP, Positive, Negative のいずれにおいても前半部に出現するケースが多く、主張における前提の共有の役割の担っている
- Value と Rhetorical Statement の出現位置の分布は OP, Positive, Negative のいずれにおいても大きな差はない
- Policy は、Positive の投稿において後半部に多く出現しており、後半における行動方針や指針の提示は説得に有効である

比較結果より、Policy を除く 4 種類では、OP, Positive, Negative の各投稿における位置に大きな差はなかったため、それらの要素をもとにどのように文章を組み立てているのかが重要となる。

5.3 議論要素の出現パターン

各投稿における議論要素の出現パターンの割合を示す。文書中の特定要素の出現パターンの計算は Wachsmuth らの提案したカウント手法を用いる [15]。

要素の出現パターンは表 2 のとおりである。なお、文中に Value の要素が多く、議論要素の位置の解析より、どの投稿においても一様に分布していたため Value を除いた出現パターンを解析するものとする。また、5%を下回るものは頻出でない出現パターンであるとして表からは除いている。

表 2 より、OP, Positive, Negative のいずれにおいても Rhetorical Statement が独立して用いられるケースが多くなっている。Positive と Negative を比較すると、Positive では Fact/Testimony の事実性の記述をした後に修辞法を用いた記述をするケースが約 15%となっているのに対して、Negative では 6.8%のみである。これにより、Rhetorical Statement は独立して述べるのではなく、事実性の記述の後

表 2 議論要素の出現パターンの頻度

	rank	Flow	%
OP	1	R	15.2
	2	T → R	12.1
	2	T	12.1
	4	R → T	6.1
	4	F	6.1
	4	P	6.1
	4	T → R → T	6.1
	4	R → F	6.1
	4	T → P	6.1
	4	F → R	6.1
Positive	1	R	28.3
	2	F → R	8.7
	3	T → R	6.5
Negative	1	R	31.8
	2	P	9.1
	3	T → R	6.8

に用いることでより説得性を高めることができる可能性がある。

また、Negative には Policy が独立して出現するケースが 9.1%あるが、Positive では 5%を下回る結果となった。議論要素の位置の解析結果を踏まえると、Positive の投稿においては後半部に Policy が集中しており、様々な要素の流れの中で Policy が最後に用いられるケースが多い。よって、主張に至る根拠を明確に示した上で行動方針や指針を提示することによって、効果的に説得を構築できる可能性がある。

6. まとめ

本論文では、説得議論フォーラムにおける投稿の特性を捉えた高粒度なアノテーションスキームの提案を行った。提案スキームにおいては、文章の意味的特性を捉えるために *Fact*, *Testimony*, *Value*, *Policy*, *Rhetorical Statement* の 5 つの要素の種類を用い、投稿内関係と投稿間インタラクションにおいては *Support*, *Attack* の 2 種類の関係を持つものとした。

また、提案スキームに基づき ChangeMyView のデータセットを無作為に 50 スレッド抽出し、議論要素の種類のアノテーションを実施した。作成したコーパスを用いて説得に効果的な特徴の解析を行った。要素の種類においては、説得成功した投稿とそうでない投稿の間には、事実文、意見文、修辞技法を用いた記述の数に違いはなく、「どの位置で述べるのか」「どういった流れで述べるのか」に違いが生じることがわかった。さらに、要素の出現パターンの解析の結果、「事実を根拠として明示した上で、行動方針や指針を提示すること」が説得に効果的である可能性があることが明らかとなった。

今後の展望としては、提案スキームを用いたコーパスの

さらなる拡張が挙げられる。また、本論文においては文章の意味的特性を捉えた要素の種類についての解析を行ったが、要素間関係や投稿間インタラクションを踏まえたコーパスの解析が期待される。さらに、解析に加えて、コーパスの機械学習の訓練データとしての利用も挙げられる。例えば、説得性の高い特徴素を用いて説得文書の判定器を作ることや、投稿文書中の要素の自動判定をするなどが考えられる。

謝辞 本研究は CREST,JST (JPMJCR15E1) および JST AIP-PRISM (JPMJCR18ZL) の支援を受けたものである。

参考文献

- [1] Cabrio, E. and Villata, S.: Combining Textual Entailment and Argumentation Theory for Supporting Online Debates Interactions, *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, Association for Computational Linguistics, pp. 208–212 (online), available from <http://aclweb.org/anthology/P12-2041> (2012).
- [2] Hidey, C., Musi, E., Hwang, A., Muresan, S. and McKeown, K.: Analyzing the Semantic Types of Claims and Premises in an Online Persuasive Forum, *Proceedings of the 4th Workshop on Argument Mining*, Association for Computational Linguistics, pp. 11–21 (online), DOI: 10.18653/v1/W17-5102 (2017).
- [3] Kirschner, C., Eckle-Kohler, J. and Gurevych, I.: Linking the Thoughts: Analysis of Argumentation Structures in Scientific Publications, *ArgMining@HLT-NAACL* (2015).
- [4] krippendorff, K.: Measuring the Reliability of Qualitative Text Analysis Data, *Quality and Quantity*, Vol. 38, No. 6, pp. 787–800 (online), DOI: 10.1007/s11135-004-8107-7 (2004).
- [5] Lippi, M. and Torroni, P.: Argumentation Mining: State of the Art and Emerging Trends, *ACM Trans. Internet Technol.*, Vol. 16, No. 2, pp. 10:1–10:25 (online), DOI: 10.1145/2850417 (2016).
- [6] Palau, R. M. and Moens, M.-F.: Argumentation Mining: The Detection, Classification and Structure of Arguments in Text, *Proceedings of the 12th International Conference on Artificial Intelligence and Law, ICAIL '09*, New York, NY, USA, ACM, pp. 98–107 (online), DOI: 10.1145/1568234.1568246 (2009).
- [7] Park, J., Blake, C. and Cardie, C.: Toward Machine-assisted Participation in eRulemaking: An Argumentation Model of Evaluability, *Proceedings of the 15th International Conference on Artificial Intelligence and Law, ICAIL '15*, New York, NY, USA, ACM, pp. 206–210 (online), DOI: 10.1145/2746090.2746118 (2015).
- [8] Park, J. and Cardie, C.: A Corpus of eRulemaking User Comments for Measuring Evaluability of Arguments, *LREC* (2018).
- [9] Peldszus, A. and Stede, M.: From Argument Diagrams to Argumentation Mining in Texts: A Survey, *Int. J. Cogn. Inform. Nat. Intell.*, Vol. 7, No. 1, pp. 1–31 (online), DOI: 10.4018/jcini.2013010101 (2013).
- [10] Stab, C. and Gurevych, I.: Annotating Argument Components and Relations in Persuasive Essays, *COLING*

- 2014, *25th International Conference on Computational Linguistics, Proceedings of the Conference: Technical Papers, August 23-29, 2014, Dublin, Ireland*, pp. 1501–1510 (2014).
- [11] Stab, C. and Gurevych, I.: Parsing Argumentation Structures in Persuasive Essays, *CoRR*, Vol. abs/1604.07370 (online), available from <http://arxiv.org/abs/1604.07370> (2016).
- [12] Stab, C. and Gurevych, I.: Recognizing Insufficiently Supported Arguments in Argumentative Essays, *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, Association for Computational Linguistics, pp. 980–990 (online), available from <http://aclweb.org/anthology/E17-1092> (2017).
- [13] Tan, C., Niculae, V., Danescu-Niculescu-Mizil, C. and Lee, L.: Winning Arguments: Interaction Dynamics and Persuasion Strategies in Good-faith Online Discussions, *CoRR*, Vol. abs/1602.01103 (online), available from <http://arxiv.org/abs/1602.01103> (2016).
- [14] Toulmin, S. E.: *The Uses of Argument*, Cambridge University Press (1958).
- [15] Wachsmuth, H. and Stein, B.: A Universal Model for Discourse-Level Argumentation Analysis, *ACM Trans. Internet Technol.*, Vol. 17, No. 3, pp. 28:1–28:24 (online), DOI: 10.1145/2957757 (2017).
- [16] Wei, Z., Liu, Y. and Li, Y.: Is This Post Persuasive? Ranking Argumentative Comments in Online Forum, *ACL* (2016).