

多人数会話における議論支援のための状態把握

高津 弘明 小林 哲則

早稲田大学 理工学術院

takatsu@pcl.cs.waseda.ac.jp

koba@waseda.jp

1 はじめに

多人数で行う議論に参加するには、議論状態—各話者の論題に対する賛成・反対の立場、話者間の協調・対立関係、賛成側と反対側での主張の優劣—を把握する必要がある。ここでは、ロボットに多人数での議論を活性化させる機能を担わせることを目的として、議論の状態を把握する方法について検討する。

議論は、主張が異なる者どうしが意見を交わすことで、自分と異なる意見の存在を理解し、知見を深める極めて知的な行為である。また、議論は、会話のあらゆる場面で生じる日常的な行為でもある。そのため、これに参加できるロボットや会話システムを作ることは、学術的にも実用的にも価値が高い。

このような背景に加え、言語処理技術や音声技術など周辺技術が発展したことで、議論の機械処理に関する研究が行われるようになってきた。日立製作所では、ビジネスや政治などの多くの場面で行われている意思決定のエッセンスが議論の一形態である競技ディベートに集約されているとして、競技ディベートで人間に勝つことを目標に、人と議論が行えるシステム(ディベート人工知能)の開発が進められている [1-6]。

これらの研究は、人と直接議論をするシステムを開発しようとしているのに対し、我々は、人同士が多人数で行う議論に参加し、これを整理・調整しながら議論の進行を支援するシステムの開発を行っている。多人数会話における議論では、競技ディベートのような形式に従った議論と異なり、話者の賛否は暗黙的であり、論題に賛成な話者と反対な話者のグルーピングも行われていない。そのため、多人数会話における議論を支援するシステムでは、このような議論の状態を把握することが必要不可欠である。

そこで、本研究では、各話者の主張が論題に対して賛成か反対か、誰と誰の意見が協調しており誰と誰の意見が対立しているか、賛成側と反対側でどちらの主張が優勢かという状態を推定する方法について検討する。

以下、2章でシステムの概要を述べ、3章で議論のための知識獲得について説明する。4章で議論状態の把握を行うシステムのディベート管理部について説明し、5章で獲得した知識を利用した発話生成について述べる。

2 システム概要

ディベートでは、論題を肯定または否定した際に生じるメリットとデメリットを議論する。そのため、各話者の発言が論題のメリットを述べているかデメリットを述べているかを解析する必要がある。ここで、メリットとは正の価値の促進または負の価値の抑制を表し、デメリットとは正の価値の抑制または負の価値の促進を表すものとする。メリット/デメリットの解析では、発話または文に含まれる促進/抑制表現を特定し、その目的語

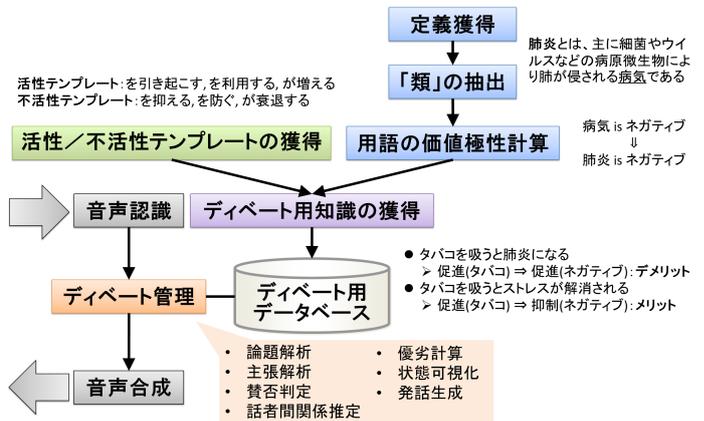


図 1: システム概要

が正の価値を持つか負の価値を持つかを判定する。そのため、促進/抑制表現に関する知識と用語の価値極性に関する知識が必要となる。促進/抑制表現は、橋本らの手法 [8] により活性/不活性テンプレートとして獲得する。用語の価値極性は Web コーパスから獲得した定義文からさらに類(上位概念)を抽出し、用語と類の関係を利用して、小規模の典型的な用語に対して人手で付与した価値極性値から膨大な用語の価値極性値を算出する。そして、これらの情報を用いて、ある概念の促進/抑制がメリットにつながるかデメリットにつながるかを解析する。この解析を Web コーパス上の文に対して適用し、促進/抑制関係で抽象化した文のデータベースを構築し、システムの発話生成で利用する。この一連の処理の流れを図 1 に示す。

3 議論のための知識獲得

3.1 活性/不活性テンプレートの獲得

促進/抑制表現は、橋本らの手法 [8] により活性/不活性テンプレートとして獲得する。ここで、テンプレートとは「助詞+用言」(e.g. が起きる, を防ぐ) からなるものを表す。また、活性とはテンプレートの項が指す対象の機能/効果/目的/役割/影響が促進・活性化されることを意味し、不活性とはそれらが抑制・不活性化されることを意味する。活性値は [0,1] の値を取り、不活性値は [-1,0] の値を取る。これらの内、絶対値が小さいものを中立とする。また、テンプレートが否定形である場合、その極性は元のテンプレートの極性を反転させたものとする (e.g. を使う [活性] → を使わない [不活性])。

まず、シードテンプレートを用意する。今回は、活性テンプレート (e.g. を引き起こす, を利用する, を高める) と不活性テンプレート (e.g. を抑制する, を防ぐ, が衰退する) を 20 個ずつ用意した。そして、シードテンプレートがコーパス上で順接か逆接を介して共起しているかどうかで項にとる名詞のペアを正の関係または負の

表 1: 名詞対の制約 表 2: テンプレート対の制約

	極性同一	極性反対		正の関係	負の関係
順接	正の関係	負の関係	順接	極性同一	極性反対
逆接	負の関係	正の関係	逆接	極性反対	極性同一

関係として抽出する(表 1)。次に、抽出した名詞対が順接か逆接を介して共起しているかどうかでテンプレート対を極性同一または極性反対として抽出する(表 2)。そして、テンプレートを電子、活性/不活性極性を電子スピンのアップ/ダウンに見立てて、ネットワークを構築し、そのネットワークのエネルギー関数を最小化することで活性値/不活性値を求める。この活性値/不活性値の絶対値が大きい上位いくつかのテンプレートをシードテンプレートに加えて再計算することでブートストラップ的に活性/不活性テンプレートを獲得する。

コーパスとして『日本語ウェブコーパス 2010』を使用し、その Web コーパスから合計 8232 個の活性/不活性テンプレートを獲得した。この内、人手で確認した 678 個の活性/不活性テンプレートをメリット/デメリット解析で使用する。

3.2 定義獲得

まず、「NP とは」とのパターンマッチングで Web コーパスから 23,004,853 個の定義文候補を抽出した。「NP とは」で始まる文が必ずしも定義文とは限らないため、識別器で定義文か否かを識別する。Web コーパスから抽出した定義文候補からランダムに 4,000 個取り出しラベルを振り(正例:604 個, 負例:3396 個)、正例に Wikipedia において「NP とは」で定義づけられる定義文 3,000 個を追加したものをデータセットとする。

非定義文の特徴は文末と「NP とは」の直後に現れやすいことから、文末の 5 形態素の単語属性とその 2-gram, 3-gram、および、「NP とは」直後の 5 形態素の単語属性とその 2-gram, 3-gram を素性として利用した。ここで、単語属性とは形態素の { 表層形, 基本形, 品詞大分類, 品詞細分類 } を表す。識別器として RandomForest と線形カーネル SVM を使用し、10 分割交差検定で評価した。なお、RandomForest の決定木の数は 200 個に設定し、乱雑さの評価基準としてジニ不純度を使用した。また、SVM のコストパラメータは 1.0 に設定した。

実験結果を表 3, 4 に示す。定義を定義と判定する精度においては、RandomForest の方が良い結果を示したが、非定義を非定義と判定する精度も考慮した場合、平均精度は線形カーネル SVM の方が良いため、定義獲得では線形カーネル SVM を使用した。この識別器を Web コーパスから抽出した 23,004,853 個の定義文候補に対して適用することで、重複なしで 5,327,918 個の定義文を獲得した。なお、用語の異なり数は 860,809 個である。

3.3 「類」の抽出

定義文はしばしば最近類 (genus proximum) と種差 (differentia specifica) によって定義付けられる。最近類とは対象を種として含む類 (上位概念) のことであり、種差とは対象を他の種から区別する特徴である。例えば、「人間とは理性的な動物である」という文において、「動物」が最近類であり、「理性的な」が種差である。

ここでは、前章で獲得した定義文から類を抽出することが目的である。類の抽出には CRF を使用した。素性に

表 3: 定義識別結果 (RandomForest)

	定義	非定義	精度	データ数
定義	3432	172	95.2%	3604
非定義	353	3043	89.6%	3396

表 4: 定義識別結果 (線形カーネル SVM)

	定義	非定義	精度	データ数
定義	3419	185	94.9%	3604
非定義	207	3189	93.9%	3396

表 5: CRF による類の抽出結果

	精度	再現率	F 値	ラベル数
I	95.9	76.5	85.1	1842
O	97.5	99.4	98.5	34113
B	94.2	82.4	87.9	1183
E	94.3	83.2	88.4	1183
S	62.5	83.3	71.4	112
平均	88.9	85.0	86.3	

は、前後 K 個の文字属性の N -gram ($K = 5, N = 1, 2, 3$) を用いた。ここで、文字属性とは { 文字の表層形, 品詞大分類の IOBES タグ, 品詞細分類の IOBES タグ, 字種 (ひらがな, カタカナ, 数字, アルファベット, 記号, その他), 対象文字が用語に含まれているかどうか } を表す。

チャンキングのラベルとして IOBES タグを定め、10 分割交差検定で評価した結果を表 5 に示す。F 値の平均は 86% 程度であった。この CRF を獲得した定義文に対して適用し、各用語の類を抽出した。その結果、例えば「急性肝炎」の類として { 肝障害, 肝臓病, 肝臓障害, 肝機能障害, 肝炎, 病気, ... } などが抽出された。

3.4 用語の価値極性計算

前章で抽出した類を利用して各用語に価値極性値を付与する。まず、いくつかの典型的な用語に対して人手で極性値を付与した。この際、価値極性が肯定であるか否定であるかは用語が活性/不活性テンプレートを伴ったときに、メリットとなるかデメリットとなるかで判断した。肯定的な価値極性を持つ用語 (e.g. 健康, 医療, 福祉) を 50 個、否定的な価値極性を持つ用語 (e.g. 病気, 感染症, 犯罪) を 50 個用意し、以下の計算を複数回 (10 回) 繰り返すことで、用語 w の価値極性値 $p(w)$ を求めた。

$$p(w) = \frac{\sum_{g \in G(w)} f_w(g) \cdot p(g)}{\sum_{g \in G(w)} f_w(g)} \quad (1)$$

ここで、 g は用語 w の類を表し、 $f_w(g)$ は用語 w における類 g の頻度を表す。ここで計算されなかった未知の用語に関しては、その用語が複数の単語から構成されている場合、構成単語の極性値の平均をその用語の極性値として与える。

3.5 デイベート用知識の獲得

Web コーパス上の文に関して、ある概念の促進/抑制がメリットにつながるかデメリットにつながるかを解析し、その情報をデータベースに保存する。例えば、「適度にお酒を飲むことは健康に良いとされる」という文があったとき、「お酒」を促進 (“を飲む”) すると、ポジティブ (“健康”) が促進 (“に良い”) されることから、この文

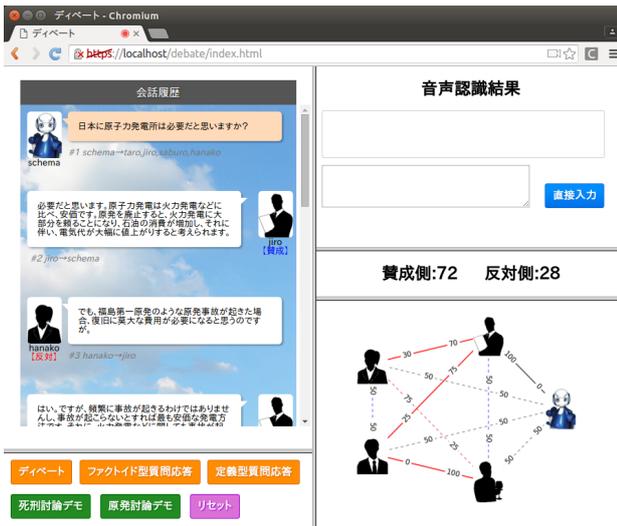


図 2: インターフェース画面

は「お酒」のメリットについて言及した文であると判断できる。一方で、「タバコを吸うと肺炎になる」という文は、「タバコ」を促進(“を吸う”)すると、ネガティブ(“肺炎”)が促進(“になる”)されることを述べており、“タバコ”のデメリットについて言及した文であると判断できる。

4 デイベート管理

4.1 論題解析

論題には、価値論題や事実論題もあるが、ここでは、政策論題を取り上げる。政策論題は、ある政策を実施すべきか否かについて議論するもので、例えば「日本は死刑制度を廃止すべきだ」や「日本政府は首相公選制を導入すべきだ」のような論題がこれにあたる。

本システムでは、まず最初にユーザーが「日本は死刑制度を廃止すべきだ」のような論題を与える。これを受け、システムは発言から活性／不活性テンプレート(“を廃止する”=抑制)を特定し、論題のテーマとなる名詞句(“死刑制度”)を抽出する。それ以外の名詞句(“日本”)は文脈として抽出する。

4.2 話者間関係推定

同じ立場の話者間の関係を協調関係と呼び、異なる立場の話者間の関係を対立関係と呼ぶ。ここでの目的は話者間が協調関係にあるか対立関係にあるかを推定することである。

基本的には表層表現に基づいて推定する。「でも」や「ですが」のような逆接で話し始めている場合、対立する主張になりやすいと考えられ、「うん」や「ですよ」のような共感・同意表現は協調を表すと考えられる。また、話者の発言の中に、相手の発言と矛盾する内容がある場合、対立した主張だと判断できる。矛盾かどうかの判定は共通の名詞(同義語を含む)を項にとる活性／不活性テンプレートの比較で行う(e.g.「死刑制度を廃止すべき」と「死刑制度を続けるべき」)。現在は、逆接や同意、矛盾の数から多数決で協調関係か対立関係を決めている。

しかしながら、この手法は直接的にコミュニケーションが成立した話者間のみでしか使えない。そこで、話者間の間接的なつながりを利用した方法も導入する。まず、全話者をリンクで結び、ネットワークを作成する。

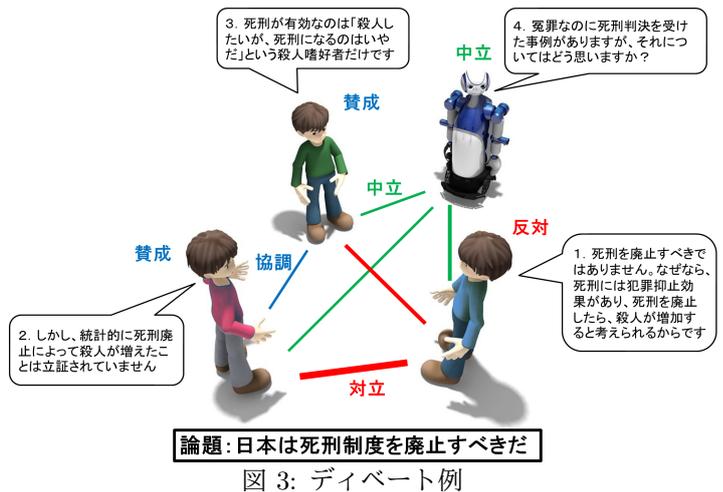


図 3: デイベート例

そして、話者 A,B,C において、A と B が対立で B と C が協調なら A と C は対立関係、A と B が対立で B と C も対立なら A と C は協調関係、というように間接的なつながりを利用した推定も行っている。

4.3 賛否判定

「死刑を廃止すべきだ」と直接的な表現で発言していれば、その内容と論題を照らしあわせて賛成か反対か特定することができる。しかし、会話形式の議論では、必ずしも自分の立場を明確に伝える義務はなく、発言の内容で暗黙的に自分の立場を示すことが多い。そこで、話者が論題に対するメリットを述べているか、デメリットを述べているかという情報を利用する。話者が論題のメリットを述べていれば、その話者は論題に賛成の立場を示していると考えることができ、逆に、論題のデメリットを述べていれば、反対の立場を示していると考えることができる。例えば、「死刑制度を廃止すると殺人が増加します」(抑制(“死刑制度”)⇒促進(“殺人”))という発言は、論題を採用したときのデメリットを述べていると言えるので、この話者は論題に反対の立場だと推測できる。この方法とは別に、協調・対立関係を利用して推定することもできる。例えば、賛成の話者と対立関係にある場合、当該話者は反対の立場だと推測できる。

現在は、論題のメリットとデメリットを述べた回数、および、対立・協調関係から推定される賛成・反対の頻度を算出し、これらのスコアを合わせて、合計が大きい方の立場をとるようにしている。

4.4 優劣計算

本システムは、特に競技デイベートでの利用を意識しているわけではないが、ロボットの参加基準に使う情報として、話者間の優劣および賛成・反対のグループ間の優劣を採点している。採点項目には「切り返し」と「論理性」、「内容」の3つの項目を定めた。切り返しは言われっぱなしにならず、きちんと反論できているかどうかに関する評価項目で、論理性は「つまり」や「したがって」などの接続詞を用いて、順序立てた説明ができていくかどうかに関する評価項目である。また、内容はメリットやデメリットに関する言及や矛盾の指摘を行っている場合、加算される。話者間の優劣スコアは、各話者のスコアをそれぞれの話者の合計スコアで割った値で与える。賛成側・反対側のグループ間の優劣スコアは、各グループに属する話者の合計スコアを両グループのスコアの和で割った値で与える。

4.5 議論状態の可視化

システムは会話の履歴および誰に向けて話したか、話者の論題に対する賛否、話者間の協調/対立関係、コミュニケーションの活性度合い、優劣スコアをインターフェース画面に表示し、議論状態の把握を支援する。実際のインターフェース画面を図2に示す。また、現実空間での議論の様子を図3に示す。

会話履歴はシーケンスに表示する。また、発言者がすぐに特定できるように発話文と一緒に話者の画像を表示している。画像の下に、話者の賛否を記し、発話文の下に誰に向けて話したかという情報を記している。また、全ての話者画像をリンクで繋ぎ、そのリンクの色で協調関係(青)/対立関係(赤)を区別し、コミュニケーションが活発な話者間のリンクがインタラクションの回数に応じて太くなるようにしている。さらに、リンクには話者間の優劣スコアを記し、賛成側・反対側のグループ間の優劣スコアをネットワークの上側に表示している。

5 発話生成

本研究では、ロボットを人間同士の議論に参加させ、議論を活性化させることを目指している。ロボットは、中立的な立場で肯定側と否定側に対して平等に意見を述べたり、劣勢(優劣スコアが低い)のグループに加勢し、立場を定めて発言したりと様々な参加方法が考えられるが、いずれの場合でも基本となるのが発話の生成である。ここでは、一人のユーザーとの議論を想定し、システムはユーザーと異なる立場で発言する場合を考える。

ユーザー発話が促進・抑制関係で抽象化できる場合、データベース上に矛盾する知識があれば、矛盾を指摘する。例えば、ユーザーが「死刑を廃止すると殺人が増えます」(抑制(“死刑”)⇒促進(“殺人”))と発言したとき、データベース上に「統計的に死刑廃止によって殺人が増えたことは立証されていません」(抑制(“死刑”)⇒促進(“殺人”)の否定=抑制(“殺人”))という知識が存在する場合、ユーザーの発言に矛盾する情報としてこれを提示する。

ユーザー発話が促進・抑制関係で抽象化できない場合、または矛盾が見つからなかった場合、ユーザーが論題に賛成なら、システムは論題のデメリットを述べる。逆に、ユーザーが論題に反対なら、システムは論題のメリットを述べる。この際、複数の候補がある場合、ユーザー発話と類似度が高く、これまでのシステム発話と類似度が低い文を選択する。類似度はBag-of-Wordsのコサイン類似度で計算している。

システムとの対話例を表6に示す。なお、音声認識にはGoogleのWeb Speech APIを使用しており、音声合成にはOpenJTalkを使用している。

6 おわりに

議論の解析に必要となる知識の獲得を行い、これらの知識を用いて議論の解析を行った。議論状態はシステムの中で管理されるが、それらの情報は参加者の意思決定に有益な情報であると考えられるため、議論状態の可視化インタフェースを開発し、議論の進行に伴う状態の変化を視覚的に捉えることを可能にした。さらに、ロボットの議論への参加で基本となるシステム発話の生成についても検討した。

表 6: システムとの対話例

ユーザー:	死刑を廃止すべきです
システム:	死刑を廃止したら殺人が増えると思う
ユーザー:	冤罪で死刑になった事例があります
システム:	ある意味では一人の人間が冤罪で死刑になる事は、国にとっては痛手では無いのだろうし
ユーザー:	少しでも冤罪の可能性があれば死刑を廃止すべきです
システム:	で、死刑を廃止すると、なにか国民の利益になるような事があるのか
ユーザー:	国家による殺人を認めてはならない
システム:	死刑を廃止したら最高刑が無期懲役になると思っています
ユーザー:	終身刑を導入すればいいと思います
システム:	日本では終身刑を導入しても死刑は廃止になりません
ユーザー:	世界は死刑廃止の流れですが
システム:	それどころか死刑を廃止すればもっと悲惨な犯罪が起きるかもしれない

現在のシステム発話は一問一答形式の単純なものに過ぎず、そのため、主張に一貫性が感じられないという問題がある。今後は、論点を分類し、発話計画を立て、一貫性のある発言や多様な観点からの発言が行えるような仕組みを考える。また、コーパス上の文をそのまま述べているため、言葉遣いが統一されていないという問題や文章が長すぎるという問題もある。そのため、発話文の要約 [14] や口語化 [15] についても検討する。

参考文献

- [1] 柳井孝介, 柳瀬利彦, 三好利昇, 丹羽芳樹, 佐藤美沙: “ディベート人工知能のためのアーキテクチャ”, 人工知能学会合同研究会 2014, 第7回データ指向構成マイニングとシミュレーション研究会, 2014.
- [2] 柳井孝介, 三好利昇, 柳瀬利彦, 佐藤美沙, 丹羽芳樹, Reisert Paul, 乾健太郎: “ディベート人工知能における意見生成”, 人工知能学会第 29 回全国大会, 2015.
- [3] P.Reisert, N.Inoue, K.Inui, T.Yanase, and K.Yanai: “Computationalizing a Toulmin Model for Argumentation Generation”, 言語処理学会第 21 回年次大会論文集, pp.828-831, 2015.
- [4] 三好利昇, 佐藤美沙, 柳井孝介, 柳瀬利彦, 丹羽芳樹: “ディベートでの立論材料文検索における論点選択方法”, 人工知能学会合同研究会 2014, 第7回データ指向構成マイニングとシミュレーション研究会, 2014.
- [5] 柳瀬利彦, 三好利昇, 柳井孝介, 岩山真, 丹羽芳樹: “ディベートの意見文章生成のための分散表現を用いた文の並び替え”, 人工知能学会合同研究会 2014, 第7回データ指向構成マイニングとシミュレーション研究会, 2014.
- [6] 佐藤美沙, 柳井孝介, 三好利昇, 柳瀬利彦, 丹羽芳樹: “議論文生成における文抽象化のための固有表現抽象化”, 第 21 回年次大会発表論文集, pp.202-205, 2015.
- [7] 木村泰知, 草原史敏, アルノゾルダン, 荒木健治: “地方議会会議録に含まれる地域課題を論題としたディベート型対話システムの提案”, 言語処理学会第 21 回年次大会論文集, pp.265-268, 2015.
- [8] C.Hashimoto, K.Torisawa, S.D.Saeger, J.H.Oh, and J.Kazama: “Excitatory or Inhibitory: A New Semantic Orientation Extracts Contradiction and Causality from the Web”, in Proceedings of EMNLP-CoNLL 2012: Conference on Empirical Methods in Natural Language Processing and Natural Language Learning, Poster, pp.619-630, 2012.
- [9] C.Hashimoto, K.Torisawa, S.D.Saeger, J.Kazama, and S.Kurohashi: “Extracting Paraphrases from Definition Sentences on the Web”, in Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, pp.1087-1097, 2011.
- [10] A.Murakami, R.Raymond: “Support or Oppose? Classifying Positions in Online Debates from Reply Activities and Opinion Expressions”, in Proceedings of the 23rd International Conference on Computational Linguistics, pp.869-875, 2010.
- [11] K.S.Hasan, and V.Ng: “Why are You Taking this Stance? Identifying and Classifying Reasons in Ideological Debates”, in Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, pp.751-762, 2014.
- [12] S.Ranade, R.Sangal, and R.Mamidi: “Stance Classification in Online Debates by Recognizing Users’ Intentions”, in Proceedings of the SIGDIAL 2013 Conference, pp.6169, 2013.
- [13] 高津弘明, 小林哲則: “対話エージェントのための性格モデル”, 言語処理学会第 21 回年次大会発表論文集, pp.191-194, 2015.
- [14] 高津弘明, 福岡維新, 藤江真也, 林良彦, 小林哲則: “会話によるニュース記事伝達のための情報選択”, 言語処理学会第 22 回年次大会発表論文集, 2016.
- [15] 高津弘明, 福岡維新, 藤江真也, 林良彦, 小林哲則: “会話によるニュース記事伝達のための口語化における述語の書き換え”, 言語処理学会第 22 回年次大会発表論文集, 2016.
- [16] 高津弘明, 福岡維新, 藤江真也, 林良彦, 小林哲則: “快適な情報享受を可能とする音声対話システム”, 言語処理学会第 22 回年次大会発表論文集, 2016.