

# 係り受け構造を用いた Web 議論掲示板における 投稿への自動分類

岩佐 幸翠<sup>1,a)</sup> 藤田 桂英<sup>1,2,b)</sup>

**概要:** 近年, タウンミーティングのように時間的, 物理的な制約に囚われる事がない, Web 上での大規模な議論のしくみが求められている. 大規模な議論のしくみを実現するために, 膨大な意見集約, 合意形成支援などファシリテートが重要となる. しかし, 参加者数や投稿数が増えるにつれ, ファシリテータが各参加者が議論の概観や議論構造を短時間で把握することは難しくなる. そこで, 我々は参加者からの投稿に対して質問, 意見, 情報, 同意, 反対, 経験といった分類を自動的に行い, 議論の構造化支援する手法を提案する. 本論文の提案手法では, 係り受け木の刈り込みにより短縮された各投稿の文, および各形態素に対して係り受け木における深さを付与された文から各形態素 N-gram の出現回数を求め, これを素性として各ラベルごとにランダムフォレストによって学習, 分類する. さらに, 実際の大規模議論データに対して人手でアノテーションしたものを正解データとして, 分類の精度などの評価を実施する.

**キーワード:** 大規模議論掲示板, 自動分類, 係り受け構造

## Automatic labeling to posts in Web discussion bulletin board using dependency structures

IWASA KOSUI<sup>1,a)</sup> FUJITA KATSUhide<sup>1,2,b)</sup>

**Abstract:** Web based Large-scale discussion platform not depending on time and physical constraints like the town meeting is required recently. In order to realize a large-scale discussion, facilitators that collect huge opinions and lead the consensus are important. However, as the number of participants and postings increases, it is difficult for facilitators to grasp the overview of the discussion and the discussion structure in a short time. Therefore, we propose a method of labelling the posts such as questions, opinions, information, favor, opposition, and experiences. In the proposed method, we extract dependencies of sentences for each sentence by using dependency analysis, and find morpheme N-gram of Bag-of-words from the end of a sentence. After that, our proposed method employs random-forests method using their features to classify the labels. Furthermore, we evaluate the precisions, recalls and f-measures of classification by using manually annotated data on actual large-scale discussion corpus.

**Keywords:** Large-scale discussion bulletin board, Automatic labeling, dependency structures

<sup>1</sup> 東京農工大学工学部情報工学科  
Department of Computer and Information Sciences, Faculty  
of Engineering, Tokyo University of Agriculture and Tech-  
nology

<sup>2</sup> 東京農工大学工学部先端情報科学部門  
Division of Advanced Information Technology and Computer  
Sciences, Institute of Engineering, Tokyo University of Agri-  
culture and Technology

a) iwasa@katfujilab.tuat.ac.jp

b) katfujic@cc.tuat.ac.jp

## 1. はじめに

シャッター商店街の町おこしからリニアモーターカーの  
開発計画に至るまで, 地域住民同士が対話して合意形成す  
る場が多数存在する. そのような場では, 各地方自治体で  
は住民参加型のタウンミーティングや市民協議会によって  
解決を図ることが目的である. しかし, 住民がタウンミー

ティングや市民討議会に参加するためには、各参加者が決められた時間に決められた場所へ集まらなければならないため、参加者の人数の制限や年齢層が偏ってしまうなど多くの問題を抱えている。実例として、大阪府堺市で平成25年度に実施された人口調査 [1] および同自治体で同年に実施されたタウンミーティングでのアンケート調査 [2] によれば、同年における60歳以上の人口は全人口に対して32%であるのに対し、同年実施されたタウンミーティングでの全参加者に対する60歳以上の参加者の割合は74.6%と大きく偏りが生じている。

このような問題を解決するために、オンライン上における大規模議論システムの研究が行われている。伊藤らは、より多数で多様な市民が参加することを前提としたWebへのアクセス環境があれば時間および場所を問わずに議論に参加可能な大規模議論システムCOLLAGREEを開発している [3]。しかし、COLLAGREEへの投稿数が増加し、議論が複雑化することによって、各参加者が議論の概観を把握するには大きな精神的、時間的負担を伴ってしまう問題を抱えている。以上から、Web議論掲示板における投稿の意図を自動で分類することで議論の構造化を支援し、各参加者が容易に議論の概観を理解できるようにすることが重要となる。

議事録や意見に対する分類、分類の研究は、合意形成を目的として対話的に意見を展開するWeb議論掲示板を対象としたものが少なく、大規模なコーパスの整備も不十分なため、有効な手法が提案されていない。例えば、談話や随筆などの非対話的な意見文書への分類、構造化に関する研究としては、Stabらによる、随筆内の談話のそれぞれの文を前提、主張、主要な主張へ分類する手法 [4] や、Smithらによる、談話における文をSituation-Entityと呼ばれる10クラスに分類することの提案など [5] が対話的に意見を展開する文書を対象に研究を進めている。しかし、これらの研究は談話などの非対話的な文書を対象とした手法であり、合意を形成するために質問や同意、反発、提案などといったWeb議論掲示板のような文書を整理するためにはこれらの分類区分ではなく、直接これらの成果を大規模議論に適用することは容易ではない。

そこで、本論文では、Web議論掲示板を対象とした投稿の内容を自動で分類する手法を提案する。具体的には、各投稿に対して文ごとに賛成、反対、質問、提案、意見、情報、経験の分類を教師あり学習によって行う手法を提案する。提案手法では、教師あり学習を用いるに当たって文章の係り受け構造に着目し、(A) 係り受け木の枝刈り込みにより短縮された文における形態素 N-gram、(B) 各形態素に係り受け木の深さを付与した形態素 N-gram の二種類の方法による素性抽出を行う。また、評価実験において、提案手法がBag-of-WordsやBag-of-形態素 N-gramによる手法に対して、F値において同等またはそれらを上回ることを示す。

以下、本論文の構成を述べる。第2章では関連研究を示す。第3章では提案手法の詳細について述べ、第4章で提案手法の評価実験について示す。第5章では本論文まとめ及び今後の課題について述べる。

## 2. 関連研究

自然言語処理において、Web上における意見の自動分類は近年研究され始めている。例えば、それぞれの意見を肯定的/否定的、賛成/反対などの二極に分類する手法が提案されている。例えば、水野らにより、ある命題に対する意見をWeb上から抽出し、各意見を同意、一部同意、一部対立、対立に分類した上で各意見の根拠を抽出することで、ある命題に対する賛成意見、反対意見とその根拠の一覧「言論マップ」を生成する手法が提案されている [6]。さらに、Mohammadらによるマイクロブログ「Twitter」における投稿を、ある議論の主題に対して肯定的か否定的かを分類する手法が提案されている [7]。また、WalkerらはWeb上での政治的なディベートにおける意見を賛成、反対に分類する手法を提案している [8]。これらの成果は、議論の主題に対する世論を客観的に抽出、集約、評価する上で非常に有効であるが、対象とする議論の主題は、意見が賛成、反対のみに自動分類されるものがほとんどである。特に、議論テーマに関する意見を集約した上で合意形成を行うことを目的としているWeb議論掲示板では、ユーザからされる投稿が賛成意見、反対意見のいずれかのであるとは限らない。例えば、「地域におけるひたくり犯罪を防止するためには？」などといったテーマでの投稿としては、具体的な対策案や解決すべき問題の具体的事例の提示などが想定される。また、Web議論掲示板では、投稿された意見についての質問やそれに対する回答として提示される情報、議論を進行する上で行われる提案、議論されているテーマに関する体験談など、意見を含まない投稿も存在する。

意見を構成する文を自動分類する研究も行われている。例えば、Stabらは、随筆における論争的な談話のそれぞれの文を前提、主張、主要な主張へ分類する手法を提案している [4]。また、Smithらは文の分類区分としてSituation-Entityを提案し [5]、これらの成果に基づいてPalmerらやFriedrichらによって新たな手法が提案されている [9][10]、例えば、FriedrichらはMASCおよびWikipediaから得た英文コーパスに対して評価実験を行い、F値71.2%という結果を出した。しかし、これらの手法は非対話的な談話によって構成された文書に対する手法であり、文書の大半が意見か前提で構成される。一方、大規模議論の文書では質問、提案、同意、対立などの対話的な文を多く含むことから、これらの手法における分類区分をWeb議論掲示板へ直接適用するのは困難である。例えば、ある事柄に対して

表 1 対話的議論における投稿文の分類区分

賛成	返信先の意見に対して肯定する文。 例) そうですね, わたしも反対です。 例) 私もそう思います。
反対	返信先の意見に対して否定する文。 例) 必ずしもそうとはいえません。 例) あなたの意見には賛成しかねます。
意見	ある事柄について考えを示している文。 例) インターネットは自由であるべきです。
情報	ある事柄について考えを示さずに事実を述べている文。 例) 渋谷にはハチ公があるらしいですね。 例) おじいさんとおばあさんがおったそう。
質問	参加者に質問をしている文。 例) 渋谷にはハチ公がありますか？
提案	参加者に提案している文。 例) 具体的な解決策について話し合ってください！
経験	ある事柄について考えを示さずに経験談を述べている文。 例) 渋谷ですごい人ごみに遭遇した経験があります。

参加者自身の意見を述べている文と単にある事柄を提示している文は Web 議論掲示板において区別されるべきであるが, Situation-Entity による分類ではいずれも状態に分類される。本論文では, Web 議論掲示板の投稿への分類区分として新たに賛成, 反対, 提案, 意見, 質問, 情報, 経験といった分類区分を定義し, Web 議論掲示板におけるそれぞれの投稿に対してこれらを自動で分類する。

### 3. 提案手法

#### 3.1 提案手法の目的

本論文では, 返信構造を備えた大規模議論掲示板における投稿文を対象として, 6 種類のラベルに基づくマルチラベル分類を教師あり学習によって行う。表 1 は具体的な対話的議論における文の分類区分を示す。なお, 本論文では議論掲示板における投稿群を「議論」, 議論における主題を「議論テーマ」と呼ぶ。

#### 3.2 提案手法の概要

本論文では係り受け構造を利用して素性を特定する。以下に, 提案手法のおおまかな流れを示す。

- 学習
  - (1) 各議論掲示板における投稿文と返信構造を取得
  - (2) 投稿文に対して前処理を実施し, 文ごとに分割
  - (3) 文ごとに人手によるラベル付けを実施
  - (4) 返信先のラベルを素性として取得
  - (5) 文の特徴ベクトル化
  - (6) 出現頻度の低い素性の削除
  - (7) SMOTE[11] によるオーバーサンプリング
  - (8) 各ラベルに対応する分類器への学習
- 分類
  - (1) 各議論掲示板における投稿文と返信構造を取得
  - (2) 投稿文に対して前処理を実施し, 文ごとに分割
  - (3) 返信先のラベルを素性として取得

- (4) 文の特徴ベクトル化
- (5) SMOTE[11] によるオーバーサンプリング
- (6) 各ラベルに対応する分類器による文の分類

#### 3.3 投稿文に対する前処理と文分割

前処理として, 投稿文に対して URL および文分割に用いない記号を削除する。さらに, 括弧や括弧に括られていない部分のうち, 「!」「!」「?」「?」「.」「.」「…」が 1 つまたは連続している部分を区切りとして文分割を行う。

#### 3.4 素性の抽出

以下に示す素性抽出法から得られた素性  $\alpha$  と  $\beta$ , およびそれぞれの文の文字長を分類器への素性として与える。

##### 3.4.1 素性 $\alpha$ : 係り受け木の枝刈り込みによる文短縮

提案手法に対して用いる議論コーパスは, 第一章でも述べた通り小規模なコーパスを想定している。そのため, 例えば「愛知の産業」に関する議論では, 意見ラベル, 情報ラベルの投稿に「名古屋」や「豊田」などの主要な都市名が頻出するなど, 小規模な対話的議論コーパスにはそれぞれの議論テーマに強く関係がある単語が多く含まれる。そのため, このようなコーパスからそのまま単語を素性として抽出した上で分類器に学習させた場合, 異なる議論テーマを扱う投稿に対する汎化性能が下がってしまう。例えば, 名古屋に関するコーパスを用いて, 「東京における犯罪防止」に関する議論の投稿文に自動分類を実行しても, 各投稿に「名古屋」や「豊田」などの単語が含まれていないため, 意見ラベル, 情報ラベルの分類が十分に行えなく, 再現率が低下する。そのため, 教師データから特定の議論テーマに関する単語を取り除く必要がある。

文書分類や文書要約の手法において, 文末表現や係り受け構造に着目する手法は多く存在する。例えば, 松本らはウェブページの主観, 客観度の判定手法を提案するに当たって, 文末表現に着目することでそれぞれのページにおけるトピックに依存しにくい手法を提案している [12]。また, 石井らは文を要約する手法において, 係り受け木における各ノード (文節) の深さが, 文における各ノード (文節) に存在する単語の重要度に対して負の相関性があると仮定し, これを利用することを提案している [13]。

これらの研究成果から, 特定の議論テーマの単語が多く含まれた少量のコーパスを教師データとして用いて分類を行う上で, 係り受け構造を用いた文短縮によって文頭, 文末表現を抽出し, 教師データから特定の議論テーマに関する単語を排除する。以下に, 提案する素性抽出手法を示す。

- (1) 文から係り受け構造における根ノードからの距離がしきい値  $d$  を超えている文節を削除 (図 1)。
- (2) 文中の各形態素を基本形に変換。
- (3) 文における各形態素 N-gram の出現回数をベクトル化。(Bag-of-形態素 N-gram)

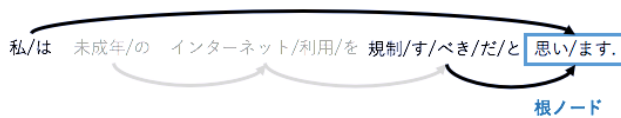


図 1 係り受け構造を利用した文短縮 ( $d = 1$  の場合)

### 3.4.2 素性 $\beta$ : 各形態素に係り受け木の深さを付与した形態素 N-gram

これまで提案されている Bag-of-Words や, Bag-of-形態素 N-gram では, 文における各形態素の位置が考慮されていないため, 賛成ラベルが付与される「その/意見/に/賛成/し/ます/.」という文における「賛成/し」という形態素 2-gram と, 賛成ラベルが付与されない「多く/の/人/が/その/法案/に/賛成/し/てい/ます/が/, /私/は/より/慎重/に/なる/必要/が/ある/と/考え/ます/. /」という文における「賛成/し」という形態素 2-gram を区別して扱うことができない. そのため, 「賛成/し」という形態素 2-gram は, 各文に対して賛成ラベルが付与されるべきか判断する上で十分な情報量を持たない. そこで, 提案する素性抽出手法では, 以下のように文から素性を得る.

- (1) それぞれの形態素に対して, 各形態素が属する分節の係り受け構造における根ノードからの距離を付与.
- (2) 文における各形態素 N-gram の出現回数をベクトル化.

### 3.5 RandomForest による学習と自動分類

提案手法では分類器として RandomForest を用いる. RandomForest とは弱学習器を決定木としたアンサンブル学習の一つであり, 分類, 回帰のために用いる機械学習アルゴリズムである [14].

### 3.6 SMOTE によるオーバーサンプリング

対話的な議論において, 表 1 に示した分類区分のうち, 多くの文は情報または意見のラベルに該当する. 一方, コーパスにおいて賛成や反対のラベルを持つ文はごく少数であり, 非常に不均衡なラベル付けがされた教師データが得られる. これらの教師データを直接 RandomForest によって自動分類を行うと, よりデータの多い分類に分類結果が偏ってしまい, 多くの賛成や反対ラベルに対応する文が正しく分類されない. そこで, 賛成や反対ラベルの教師データをオーバーサンプリングによって擬似的に増加させることで本課題を解決する. 提案手法では, オーバーサンプリングの手法として SMOTE[11] を用いる. SMOTE は, 少数派であるクラスにおける各点について, ランダムに  $k$ -最近傍点までの点を生成した上で, 多数派であるクラスに対してランダムにダウンサンプリングを行う.

## 4. 評価実験

### 4.1 実験設定

表 2 に示す 4 つの大規模議論データセット (ネットリテ

議論テーマ	投稿者数	投稿数	文の数
1 ネットリテラシーの問題点	59	283	926
2 名古屋における災害	21	332	1162
3 名古屋における魅力	34	392	1422
4 名古屋を活気づけるために	34	244	903

ラシーの問題点, 名古屋における災害, 名古屋における魅力, 名古屋を活気づけるために) を用いて, 1 つの議論データを評価データ, それ以外の 3 つの議論を教師データとして, 全結果におけるベースライン (A), ベースライン (B), 提案手法の適合率, 再現率, F 値を比較する.

### 正解データセット

表 2 に示した議論群に対して, 東京農工大学の学生 3 名による人手での評価を実施し, 多数決によってそれぞれの文への正解分類ラベルを決定した.

### 提案手法のパラメータ設定

提案手法およびベースラインが最良の結果となるように, パラメータの設定を行った. 以下に各手法のパラメータを示す. 提案手法では, 素性  $\alpha$ , 素性  $\beta$  で用いる形態素 N-gram について,  $N = 2, 3$  とする. また, 素性  $\alpha$  については  $d = 1$  とする. RandomForest におけるハイパーパラメータとして, 決定木の最大の深さ  $D$  と木の本数  $B$ , 一度の決定木生成においてランダムに選択する素性の数  $F$ , 各決定木におけるノードの最大数  $N$  がある. 本論文では, これらのパラメータをグリッドサーチによってチューニングした上でランダムフォレストを学習させた結果,  $D = \infty$ ,  $B = 1000$ ,  $F = \sqrt{FNUM}$  ( $FNUM$  は素性の数),  $N = \infty$  とした. SMOTE において参照する  $k$ -最近傍点については  $k = 5$  とした.

### ベースライン手法

評価実験において, 提案手法と比較するために, 以下をベースライン手法とした.

- ベースライン (A): Bag-of-words
- ベースライン (B): Bag-of-形態素 N-gram ( $N = 2, 3$ )

ベースライン手法 (A), (B) では, 提案手法と同様に, 素性の抽出後は出現頻度の低い素性を削除し, SMOTE によってオーバーサンプリングした上で, Random Forest によって学習, 分類をする.

### 4.2 実験結果と考察

表 3 は各手法に対する評価実験の結果をラベル別に示している. 提案, 情報, 経験ラベルでは, 提案手法が F 値においてベースラインより良い結果であった. 一方, 賛成, 質問, 意見ラベルでは提案手法がベースラインを下回っている. なお, 反対ラベルについていずれの手法でも F 値は 0% となった原因としては, 評価用データセット中に反対ラベルに分類されるような文が殆ど存在せず, 十分に学習を行うことができなかったためである.

賛成ラベルにおいて、提案手法がベースラインに対して上回る性能を発揮できない理由は、賛成ラベルには「賛成です。」「全くその通りだと思います。」「私もその考えに賛成です。」などの短く簡潔な文が多く、議論テーマに固有な単語が含まれる場合が少ないためである。提案手法における素性  $\alpha$  で用いた係り受け木の枝刈り込みによる文短縮のみでは、期待したような議論テーマに固有な単語を削除することによる汎化性能の向上効果を得ることは難しい。また、「すごく良いアイデアですね!」「良いアイデアだと思います!」などの文章における形態素「良い」や形態素 2-gram 「良い/アイデア」など、係り受け構造によらず特定のラベルに出現しやすい形態素 (または形態素列) が存在するが、提案手法では係り受け構造によらない素性について着目することができないことも理由として考えられる。以上から、提案手法に係り受け構造に着目しない素性抽出を含める必要があることが今後の課題としてあげられる。

また、質問、意見ラベルでも、提案手法がベースラインに対して上回る性能を発揮できなかった。この結果の理由として、意見、質問ラベルに分類される文が評価用データセットに十分な量が含まれていないために、ベースラインにおいて各議論に含まれる議論テーマに固有ないし強い関連のある単語による汎化性能の低下が起こらなかったことによる。

一方で、提案、情報、経験ラベルでは、提案手法はベースラインに対して F 値を 0.1 以上上回ることができた。この理由として、質問、意見ラベルにおいて提案手法の性能がベースラインを上回らなかった場合と逆の理由であり、議論テーマに固有ないし強い関連のある単語による汎化性能の低下したことに起因する。実際に、提案手法が当該ラベルにおいて、再現率がベースラインに対して上回っている。

また、ベースライン及び提案手法では、同じ単語を含んでも文脈によって意味が変わるような文へ対応することが出来ないという課題がある。例えば、同じ「そうですね。」という文でも、これは相手の意見に対する同意としての「そうですね。」と、相槌としての「そうですね」が存在する。そのため、対象となる文が「賛成」ラベルを付与されるべきかは、対象となる文自体から得られる情報では判定することが難しい。今後、対象の前後にある文の形態素や評価極性を素性として与えることが考えられる。

さらに、ベースライン及び提案手法は共に隣り合う形態素に着目する一方で、必ずしも各ラベルにおいて出現しやすい形態素のペアが隣り合うとは限らないため、十分に分類のための素性を抽出することができていない可能性がある。例えば、「すごく良いアイデアですね!」「良いアイデアだと思います!」という賛成ラベルに分類される文について考える。「良い」という形態素と「!」という形態素が、賛成ラベルに分類されるような文に出現しやすいと仮定した場合、ベースラインおよび提案手法は共にこの情報

表 3 各手法に対する評価実験の結果 (ラベル別)

手法	ラベル	適合率	再現率	F 値
ベースライン (A)	賛成	0.46	0.48	0.47
	反対	0.00	0.00	0.00
	質問	0.75	0.41	0.54
	提案	0.82	0.16	0.27
	意見	0.76	0.79	0.78
	情報	0.43	0.13	0.20
	経験	0.57	0.06	0.10
ベースライン (B)	賛成	0.60	0.48	0.53
	反対	0.00	0.00	0.00
	質問	0.78	0.68	0.72
	提案	0.67	0.25	0.36
	意見	0.80	0.77	0.79
	情報	0.55	0.15	0.24
	経験	0.62	0.18	0.28
提案手法	賛成	0.57	0.50	0.53
	反対	0.00	0.00	0.00
	質問	0.73	0.71	0.72
	提案	0.70	0.34	<b>0.46</b>
	意見	0.80	0.78	0.79
	情報	0.52	0.27	<b>0.36</b>
	経験	0.46	0.29	<b>0.36</b>

を利用しない。そのため、提案手法の今後の展望として、共起する形態素に着目するための方法を形態素 N-gram から何らかの方法へ変更する事が考えられる。

## 5. おわりに

本論文では、大規模議論掲示板における投稿の構造化を目的として、文ごとに賛成、反対、質問、提案、意見、情報、経験の分類区分に対してラベリングする手法を提案した。提案手法では、まず、投稿を文ごとに分割し、素性を抽出した上で、SMOTEによりオーバーサンプリングを行う。その後、ランダムフォレストによって学習および分類を行う。素性を選択するに当たって、議論テーマに固有な単語を排除することで汎化性能を向上させるために、係り受け木の枝刈り込みによって短縮された文、および形態素に係り受け木の深さを付与された文のそれぞれから形態素 N-gram を抽出した。提案手法は、対話的議論データを構造化する上で、小規模なコーパスから学習しなければならない場面において有効であった。また、評価実験の結果、特定のラベルに対してベースラインを上回る高い F 値であった。

今後の展望として、提案手法の性能をさらに向上させるための、返信構造や前後文を利用した素性や係り受け構造に依存しない素性の導入や共起する形態素をより広く探索できる方法の検討、導入が挙げられる。

## 参考文献

- [1] 大阪府 堺市役所 市長公室 企画部調査統計担当：堺市全市・区域別年齢別人口，大阪府堺市役所（オンライン），入手先（<http://www.city.sakai.lg.jp/shisei/tokei/nenreibetsu/zensikunenrei.html>）（参照 2016-10-31）.
- [2] 大阪府 堺市役所 市長公室 企画部調査統計担当：堺市平成 25 年度タウンミーティング，大阪府堺市役所（オンライン），入手先（[http://www.city.sakai.lg.jp/shisei/gyosei/kusei/townmeeting/townmeeting\\_h25.html](http://www.city.sakai.lg.jp/shisei/gyosei/kusei/townmeeting/townmeeting_h25.html)）（参照 2016-10-31）.
- [3] 伊藤孝行，奥村 命，伊藤孝紀，秀島栄三：多人数ワークショップのための意見集約支援システム Collagree の試作と評価実験：議論プロセスの弱い構造化による意見集約支援，日本経営工学会論文誌，Vol. 66, No. 2, pp. 83-108（オンライン），DOI: 10.11221/jima.66.83（2015）.
- [4] Stab, C. and Gurevych, I.: Identifying Argumentative Discourse Structures in Persuasive Essays., *EMNLP*, pp. 46-56（2014）.
- [5] Smith, C. S.: *Modes of discourse: The local structure of texts*, Cambridge University Press（2003）.
- [6] 水野淳太，渡邊陽太郎，エリックニコルズ，村上浩司，乾健太郎，松本裕治：文間関係認識に基づく賛成・反対意見の俯瞰，情報処理学会論文誌，Vol. 52, No. 12, pp. 3408-3422（オンライン），入手先（<http://ci.nii.ac.jp/naid/110008719918/>）（2011）.
- [7] Mohammad, S. M., Kiritchenko, S. and Zhu, X.: NRC-Canada: Building the State-of-the-Art in Sentiment Analysis of Tweets, *CoRR*, Vol. abs/1308.6242（online），available from（<http://arxiv.org/abs/1308.6242>）（2013）.
- [8] Walker, M. A., Anand, P., Abbott, R., Tree, J. E. F., Martell, C. and King, J.: That is your evidence?: Classifying stance in online political debate, *Decision Support Systems*, Vol. 53, No. 4, pp. 719-729（2012）.
- [9] Palmer, A., Ponvert, E., Baldrige, J. and Smith, C.: A sequencing model for situation entity classification, *ACL*, pp. 896-903（2007）.
- [10] Friedrich, A., Palmer, A. and Pinkal, M.: Situation entity types: automatic classification of clause-level aspect, *Proc. of ACL*（2016）.
- [11] Chawla, N. V., Bowyer, K. W., Hall, L. O. and Kegelmeyer, W. P.: SMOTE: synthetic minority over-sampling technique, *Journal of artificial intelligence research*, Vol. 16, pp. 321-357（2002）.
- [12] 松本章代，小西達裕，高木 朗，小山照夫，三宅芳雄，伊東幸宏：文末表現を利用したウェブページの主観・客観度の判定，第 1 回データ工学と情報マネジメントに関するフォーラム（DEIM），A5-4（2009）.
- [13] 石井弘志，林日華，古郡廷治：単語の中心性に基づくテキスト自動要約システム，情報処理学会研究報告自然言語処理（NL），Vol. 2001, No. 20, pp. 83-90（オンライン），入手先（<http://ci.nii.ac.jp/naid/110002934300/>）（2001）.
- [14] Breiman, L.: Random Forests, *Machine Learning*, Vol. 45, No. 1, pp. 5-32（online），DOI: 10.1023/A:1010933404324（2001）.