

議会における明示的な賛否表現を含まない議員発言を対象としたスタンス分類データセットの構築と評価

猫本隆哲¹ 秋葉友良¹ 増山 繁²

¹ 豊橋技術科学大学 ² 東京理科大学

nekomoto.takanori.pf@tut.jp akiba@cs.tut.jp masuyama@rs.tus.ac.jp

概要

スタンス分類の目的は、発信者が特定のトピックに対して賛成なのか反対なのかを識別することである。近年では、SNSにおいて様々なトピックに関する議論や討論を獲得し、スタンス分類のための教師データとして用いられている。本研究では、議会議事録に収録されている議員発言を対象としたスタンス分類を提案する。提案するスタンス分類問題に対し、議会での討論のパターンを考慮することにより、議会議事録から自動的にデータセットを構築した。また、構築したデータセットを用いて分類機を構築し、評価実験を行った。

1 はじめに

本研究では、NTCIR15 QA Lab-PoliInfo-2 Stance Classification タスク [1] から派生した文書分類タスクを提案する。QA Lab-PoliInfo-2 タスクでは、議会議事録から議員の意見、その理由・条件の要約、分類等を行い、表示するといった複雑な実世界の質問応答技術を目指している。Stance Classification では、都議会における政治家の発言から、政治家の立場(スタンス)を推定することを目的としている。より具体的には、議会議事録から、議会中に議論される各議題について、各会派の賛否を分類する問題に取り組んだ。日本の議会議事録の記述は、議員ごとの発言の書き起こしであり、ほとんどの議員は討論の冒頭で複数の議案に対して簡単な賛否を述べることから、ここから一定数の議案に対してのスタンスが判明する。Stance Classification の評価結果では、この冒頭部分のみを用いることで90%以上の正解率で賛否分類できることが示されている [2]。しかし、一部には冒頭部分で賛否表明を行わない、または冒頭のみでは賛否表明されない議案が存在するといった場合があり、そのような場合に対応するには、冒頭以外の討論部

分から賛否分類を行う必要がある。

本研究では、我々の先行研究 [3] に引き続き、ある議案に対して明示的な賛否の表現のない討論の発言から賛否分類を行う問題に焦点を当てた。議会における議員討論の特性を調査し、トピックとなる議題に対応した討論の抽出から賛否のラベリングまでを自動化し、データセットを構築した。また、構築したデータセットを用いて分類器を学習し、評価実験を行った。

2 スタンス分類

本節では、Stance Classification タスクの対象となった東京都議会を例として、議事録の構造、および本研究の問題設定について詳細を述べる。

2.1 会議録データ

都議会の討論の特徴として、議員は図 1 のように発言の冒頭で特定議題への一定の賛否を述べてから詳細な討論に入る。

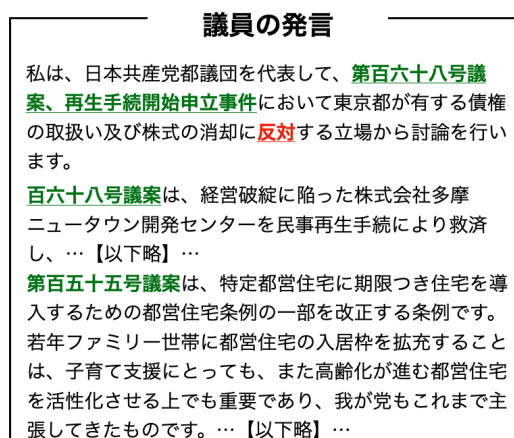


図 1 議員発言例

議員は必ず議長から発言権が与えられた上で、始めにいくつか議案への簡単な賛否表明を行った上で、それぞれの議案についての詳細な討論を行う。こ

ここでは、前者を**賛否表明発言**、後者を**討論発言**と定義する。図 1 から切り離した賛否表明発言および討論発言をそれぞれ図 2, 図 3 に示す。

賛否表明発言

私は、日本共産党都議団を代表して、**第百六十八号議案、再生手続開始申立事件**において東京都が有する債権の取扱い及び株式の消却に**反対**する立場から討論を行います。

図 2 賛否表明発言例

賛否表明発言は、議長から発言権が与えられた議員が始めに述べる形式的な賛否表明である。多くの場合、ここから特定議案（議案番号）の賛否表明が得られ、Stance Classification タスクの評価結果では、賛否表明発言のみを参照することで 90%以上の正解率が実現されることが分かっている [2]。

討論発言

百六十八号議案は、経営破綻に陥った株式会社多摩ニュータウン開発センターを民事再生手続により救済し、…【以下略】…
第百五十五号議案は、特定都営住宅に期限つき住宅を導入するための都営住宅条例の一部を改正する条例です。若年ファミリー世帯に都営住宅の入居枠を拡充することは、子育て支援にとっても、また高齢化が進む都営住宅を活性化させる上でも重要であり、我が党もこれまで主張してきたものです。…【以下略】…

図 3 討論発言例

討論発言では、賛否表明を行った議案への詳細な言及がなされる。賛否表明発言と比較すると、同一議案への言及でありながらも、それについての根拠が明確に述べられている。しかし、討論発言では賛成や反対といった単語は必ずしも出現しない。また、賛否表明発言では言及されず、その後の討論発言のみで言及される議案も存在する。また、討論発言では複数の議案について議案ごとに順番に発言が行われる。よって、発言者のある議案に対するスタンスを捉えるためには、討論発言を議案ごとに分割したテキスト（セグメント）を獲得する必要がある。

2.2 問題設定

前述したように賛否表明発言を用いることで高い精度で賛否分類を行うことが可能であるが、陽に言及されない議題には対応できない。そこで本研究では、図 3 に示した討論発言から、議案に対する賛否を分類する問題に焦点を当てる。そのために、議事録

の各議員の各議案に対する討論発言のセグメント分割を行い、その議題への賛否を対応付けたペアデータを構築した。このペアデータを学習データとして、二値分類器の構築、評価を行った。

3 データ構築手法

本節では、議会議事録からの教師データの構築方法、それに伴う討論発言のセグメント分割手法について述べる。

3.1 セグメント分割

議案に対応する討論発言を獲得するために討論のセグメント分割を行う。セグメント分割には、議員討論で多用される表現方法、特に特定の接続詞に注目して分割を行う。議事録の場合、接続詞が出現する箇所は言及する議題に変化を予測しやすいため、図 4 のような議員の討論発言例では、「次に、」という接続詞に注目して分割を行う。また接続詞以外に、議事録に記載されている議案番号や議案名にも注目し、分割を行う。注意する点としては、データセットには議題と関連の無いノイズとなるような発言の混入を防ぐために、議案番号と議案名については討論発言との完全一致を主として分割を行っていく。

しかしながら、議員が議事録記載の議題名をそのまま発言するとは限らない。そこで、可能な限り議題対応が取れる多くの発言を獲得するために、別途編集距離を用いる。手法としては、図 4 のように討論内に「～について」という発言が含まれる場合、それ以前の発言と議題名との編集距離を測る。その類似度があらかじめ定めたしきい値を超える場合、議題対応が取れる発言であるとみなし、セグメント分割を行うものとする。本研究では、しきい値を 0.9 に設定した。これらのルールベース手法で獲得したセグメントを総じて**セグメント候補**として利用する。

3.2 ラベリング

次に、獲得したセグメント候補にラベリングを施す。セグメント候補に対するラベリングは次の 3 つの手法を検討した。

議決参照 議事録の議決部から賛否を与える。

賛否表明参照 A 議決参照せず、賛否表明発言から賛否を与え、不明部分には賛成ラベルを与える

賛否表明参照 B 議決参照せず、賛否表明発言から賛否を与え、不明部分は教師データに含めない。

議員の討論文

次に、株式会社新銀行東京に関する特別委員会の継続調査について意見を述べます。そもそも都議会公明党は、本特別委員会の設置について、平成二十一年第三回定例会で異議を申し立てました。それは、特別委員会を設置しなければ質疑ができないというものではないからであり、現にあらゆる角度から本会議、予算特別委員会、経済・港湾委員会において審議を重ねてまいりました。一方、本委員会においては、設置後二年間、特別委員会ならではの活発な審議が行われたとは到底思えず、委員会を設置しておく意味はないものと考えます。なお、新銀行東京については、今後、さらに企業価値を高めた上で、事業譲渡あるいは業務提携を行い、追加出資の回収もしくは保全を行っていくべきと申し上げておきます。

次に、東京都中央卸売市場築地市場の移転・再整備に関する特別委員会の継続調査についてです。現在地再整備の可能性については、本特別委員会及び小委員会において、これまで長期間にわたり、また、休日にも審議を行うなど精力的に検討した上で、実行不可能であることが明らかになりました。都議会公明党としては、結論が明白である以上、特別委員会をこれ以上継続する必要性は全くないものと考えます。

議事録記載の議案番号、議案名

第百十一号議案:

株式会社新銀行東京に関する特別委員会の継続調査

第百二十七号議案:

東京都築地市場の移転・再整備に関する特別委員会の継続調査

→ Levenshtein . ratio(seg1,seg2) > 0.90?

図4 討論発言のセグメント分割および編集距離を用いた議題特定

東京都議会の場合、議事録に収録されている議決を参照することで各政党ごとに各議案に対する賛否が分かる。しかし、地方議会や国会などでは賛否数のみの記載に留まることがあるため、賛否表明発言からラベルを与える手法を検討した。賛否表明参照 A と B の違いは、賛否表明発言で言及されていないセグメント候補に賛成ラベルを与えてデータセットに採用するか、もしくはラベルが与えられないセグメント候補はデータセットに含めないかである。よって、賛否表明参照 B は、賛否表明参照 A よりも誤ラベル率が低い代わりに、サンプル数が少なくなる傾向がある。これが分類実験ではどのような影響を及ぼすかを検証する。

3.3 データセット総括

本手法を 2000~2020 年までの東京都議会に適用し、データセットの自動構築を行った。表 1 にラベリング手法ごとに構築されたデータセット総数を示す。

ラベリング	train	error
議決参照	539	0
賛否表明参照 A	539	12
賛否表明参照 B	500	5

構築できたデータセットは 539 件となり、推定賛成ラベルを与えたセグメント候補をデータセットに含めない場合は 500 件となった。また error は議決を参照せずにラベルを予測したうち、教師データのセ

グメント候補に対して与えられた誤ラベル数を示している。これを学習に用いて人手で構築した 137 件のテストセットを用いて評価実験を行う。

4 評価実験

本節では、構築したデータの分類に用いる分類器の概要、分類器の評価実験、考察について述べる。

4.1 BERT を用いた分類

2.1 節で定義した討論発言を対象に議員のスタンスを 2 値 (賛成 or 反対) に判定する BERT ベースの分類器を構築した [4]。モデルには、公開されている日本語事前学習済みモデルの BASE 通常版を用いた [5]。本モデルは、JUMAN [6] による単語分割に基づいて学習されているため、同じ形態素解析器を使用している。また、今回は入力が討論発言、出力を賛成または反対とすることが目的であるため、[CLS] トークンの最後の隠れ状態を入力 768 次元、出力を賛成または反対の 2 クラスとする全結合層に通して予測値を出力する。

4.2 入力構成

入力は単一討論発言、または討論発言に議題名を加えたものの 2 つを検証する。議題名を追加で与える試みについては、獲得したセグメント候補は特定の接続詞や発言に出現する議題名に注目して分割を行っている。そのため、セグメント候補には議題とは関連のないノイズが含まれることがあるため、議題

名を追加で与えることで分類精度に影響を及ぼすか検証するために行う。

4.3 実験設定

作成した3つのラベリング手法との比較対象として以下の3つのベースラインを設定した。

Majority 討論内容に関わらず全て賛成と分類する。

Baseline1 討論発言に出現する「賛成」「反対」の単語数をカウントし、多い方にラベルを与える。ただし同数、または出現しない場合は「賛成」とする。

Baseline2 知事提出議案を一括で賛成、議員提出議案を一括で反対とする。

ハイパーパラメータの設定は、入力最大の系列長を512、ミニバッチサイズを24、学習係数を $2e-5$ 、エポック数を3とした。

4.4 結果と考察

分類手法	ラベリング	付加	accuracy
Majority			0.7299
Baseline1			0.5766
Baseline2			0.7518
BERT	議決参照		0.8248
BERT	議決参照	議題	0.8467
BERT	賛否表明参照 A		0.7299
BERT	賛否表明参照 A	議題	0.8029
BERT	賛否表明参照 B		0.8029
BERT	賛否表明参照 B	議題	0.8321

表2に3つのラベリング手法それぞれに議題を与えたものの合計6種の精度を示している。まず、majorityの時点で分類精度が7割を超えており、構築されたデータセットでの議員発言はそのほとんどが賛成のスタンスであることが分かる。その他のベースラインではbaseline2が最も精度が高く、知事提出議案に対しては賛成、議員提出議案については反対のスタンスを表明する議員が多数いることが予想できる。提案手法については全手法でbaselineを上回っており、また議題を追加で与えることで精度向上に寄与していることが分かった。また、正解ラベルを議決から与えた議決参照ラベリングが最も精度が高くなり、誤ラベルが少ないtrainセットを用いた方が次いで正解率が高い。よって、今回の実験ではデータセット総数よりも、ラベリング精度を重視する方

が精度向上に好影響を与えることが分かった。

賛成 1 まず最初に、宿泊税についてです。今回、宿泊税は、法定外目的税として提案されました。私たち生活者ネットワークは、地方分権一括法が昨年四月に施行され、課税自主権が地方に付与された条件を生かそうとする試み自体は高く評価するものです。しかし、具体的提案については、率直に使い道の内容やプロセスの点で、大きな疑問を提起せざるを得ません。日本はこれまで、黒字を解消するため海外へ出ることを奨励し、観光を政策として進めてきませんでした。これは東京都も同じで、東京の過密化をさらに進める観光客の誘致をよしとはしてきませんでした。(以下略)

図5 誤分類例(誤って賛成ラベルに分類された例)

図5に実際に分類器へ討論発言を入力し、誤分類された例を示す。この発言は教師あり分類器に共通して、正解ラベルは反対でありながらも誤ラベルである賛成に分類されてしまった討論発言である。誤分類された討論には、肯定意見と否定意見が混在しており、また肯定側の討論はよく使われる「評価する」といった表現が含まれていることから、賛成と分類されたと考えられる。また、このような機械が誤分類する討論発言は、人間が確認を行っても誤分類してしまう傾向が高いため、同様の誤分類が他の議会議事録においても発生するか、今後も検証を続ける必要がある。

5 おわりに

本研究では、議会の議員発言を対象として、その特性を活用したアノテーションを不要とするデータセットの構築方法を提案した。また、BERTベースのスタンス2値分類器を構築し、構築したデータセットのベースラインを確立した。今後はより多くの議員討論を獲得するために、地方議会等を対象として更なるデータセットを構築し、提案手法の有効性を検証する。

謝辞

本研究はJSPS 科研費19K11980の助成を受けた。

参考文献

- [1] Yasutomo Kimura, Hideyuki Shibuki, Hokuto Ototake, Yuzu Uchida, Keiichi Takamaru, Madoka Ishioroshi, Teruko Mitamura, Masaharu Yoshioka, Tomoyoshi Akiba, Yasuhiro Ogawa, Minoru Sasaki, Kenichi Yokote, Tatsunori Mori, Kenji Araki, Satoshi Sekine, and Noriko Kando. Overview of the ntcir-15 qa lab-poliinfo-2 task. **Proceedings of The 15th NTCIR Conference**, 12 2020.
- [2] Takanori Nekomoto, Ryoto Ohsugi, Tomoyosi Akiba, Shigeru Masuyama, and Daiki Shirato. akbl at the ntcir-15 qa lab-poliinfo-2 tasks. **The 15th NTCIR Conference**, p. 155, 2020.
- [3] 猫本隆哲, 秋葉友良, 増山繁. 議題への明示的な賛否表現を含まない議員発言を対象とした議会議事録のスタンス分類. 言語処理学会第 27 回年次大会 発表論文集, pp. 669–673, 2021.
- [4] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. **arXiv preprint arXiv:1810.04805**, 2018.
- [5] Tomohide Shibata, Daisuke Kawahara, and Sadao Kurohashi. Improving the accuracy of japanese parsing with bert. **The Association for Natural Language Processing 25th Annual General Meeting**, pp. 205–208, 2019.
- [6] Daisuke Kawahara. Japanese morphological analysis system juman version 3.61. <http://pine.kuee.kyoto-u.ac.jp/nl-resource/juman.html>, 1999.

A 付録

本節に構築したデータセットに収録されている実際の賛成討論, 反対討論をそれぞれ示す。

賛成 1

第二百七号議案、東京都情報公開条例の一部を改正する条例によって、情報公開の実施機関に公安委員会と警視総監が加えられました。今期、全国初の都議会独自の情報公開条例が実現し、さらに今、警察情報が公開実施機関に加えられるということは、開かれた政治への前進として、心から歓迎するものです。しかしながら、この警察情報の公開が真に実効あるものとなるためには、開示義務除外とされる情報を、犯罪捜査などに支障を及ぼすおそれのある刑事法の執行を中心としたものに限定し、いわゆる行政警察活動や犯罪捜査などの個別の警察活動に支障を及ぼさない財政支出などは原則開示とするなど、明確な基準による運用が必要です。都民の期待にこたえる地方自治体としての積極的な条例運用を求めます。、小松恭子,2000/7/12, 第二百七号議案, 東京都情報公開条例の一部を改正する条例

反対 0

犯罪被害者等基本条例について申し上げます。総務委員会における都議会公明党の質疑で、本条例案の数々の問題点が露見しました。条例の根幹は目的と定義にあります。しかし、本条例の目的と定義は、国の犯罪被害者等基本法の条文の中の国を都に、また、国民を都民にほぼ置きかえただけであり、犯罪等の範囲や犯罪被害者等の対象の解釈についても国の解釈を踏襲しており、都としての独自性や必要性がないことが明らかにされました。また、条例案第十一条の経済的負担を軽減する生活資金の給付については、現在、国が具体的に見直しを検討しており、今回の質疑においても具体的な給付案及び根拠となる財源は明示されませんでした。さらに、具体的な給付案については、都知事が設置する犯罪被害者等施策推進会議でもまれるとの答弁があり、都に丸投げする形になっています。また、既に都が設置している東京都犯罪被害者等支援推進会議との関係からすれば、屋上屋を重ねるだけであることが明らかであります。公明党は四十六年前、凶悪犯罪によってかけがえないご子息を失った一人の父親の声を聞き、以来、犯罪被害者等の支援に取り組んでまいりました。今日までの半世紀にわたる取り組みは、犯罪被害者等が抱える悲しみ、苦しみに少しでも寄り添い、心を配りながら進めてきました。このたびの条例案は余りにも拙速かつ不十分な内容であり、形だけの条例案といわざるを得ず、到底賛成することはできません。、小林健二,2012/10/4, 議員提出議案第十一号, 東京都犯罪被害者等基本条例