

地方議会議事録を対象としたスタンス分類データセットの自動構築と ベースライン分類器の評価

我藤勇樹¹ 秋葉友良¹ 内田ゆず² 木村泰知³ 高丸圭一⁴

¹豊橋技術科学大学 ²北海学園大学 ³小樽商科大学 ⁴宇都宮共和大学

gato.yuki.am@tut.jp akiba@cs.tut.ac.jp yuzu@hgu.jp kimura@res.otaru-uc.ac.jp
takamaru@kyowa-u.ac.jp

概要

スタンス分類は、特定のトピックに関して述べられた意見が賛成なのか反対なのかを分類することを目的としている。本研究では、複数の地方議会議事録から自動でスタンス分類データセットを構築する手法を提案する。愛知県の種々の規模の議会を対象に本手法を適用し、6,254 サンプルのデータセットを構築することができた。構築したデータセットを手で評価したところ、抽出漏れは約 10%あったものの賛否ラベルの付け誤りは 0%であった。また、構築したデータセットを基にベースライン分類器を作成し、分類性能の評価を行った。

1 はじめに

スタンス分類は、特定のトピックに関して述べられた意見が賛成なのか反対なのかを分類することを目的としている。例えば、「マスク着用」というトピックに対する意見「周囲の人を守るためにマスクを着用します。」は、賛成のスタンスを表明していることになる。スタンス分類が自動で行われることで、ソーシャルメディアに流れる大量のテキストを分析し、世論の把握やファクトチェックに活用することができる。

2019 年後半から 2020 年にかけて開催された NTCIR15 QA Lab-PoliInfo2 では、東京都議会を対象とした Stance Classification(SC)タスク [1]が実施された。これは、会議議事録全体を参照し各会派の議案に対するスタンス(賛成/反対)を推定するタスクであった。議会議事録中には、討論の冒頭に議案に対する賛否を明示的に表明する発言があり、この部分から賛否表現を抽出するだけで 90%以上の精度を達成することができた。一方、純粹に発言の内容からスタンスを分類するという本来のスタンス分類

に対する性能は明らかになっていない。

○議長 (松下浩)

ただいまの総務委員長の報告に対する質疑に入ります。質疑はありませんか。(「なし」の声あり) ないようですので、質疑を終わります。これより討論に入ります。まず、議案第59号について討論を行います。討論はありませんか。賛成の討論、8番・今井美千代議員。

◆8番議員 (今井美千代)

議案第59号「平成14年度大府市一般会計補正予算(第1号)」について、賛成の立場で討論いたします。障害者福祉制度が来年度4月から大きく変わろうとしており、現在、障害のある方が施設やホームヘルプなどのサービスを利用する場合、措置制度によって国と自治体がサービスの提供に直接的な責任を負っていますが、来年度4月からは、障害のある方本人が利用するサービスを決定し、自らがサービス事業者を選んで契約する仕組みになります。

(省略)

最後に、愛知県に対しては直営の障害者施設を民間に移譲する動きや民間福祉施設への人件費補助を廃止する動きがあるが、万博空港に税金をつぎ込み福祉予算を削る施策に対しては厳しく批判し、直営の施設、在宅サービスの維持、民間施設への助成の充実を行うよう強く要請されること、以上を福祉の低下にならないよう提案し、私の賛成討論とさせていただきます。

図 1 実際の討論発言 (愛知県大府市議会)

本研究では、明示的な賛否表現に頼ることなく発言のスタンス(賛否)を推定する問題に取り組むために、地方議会議事録から自動的にスタンス分類データセットを構築する手法を開発し、データ構築を行った。PoliInfo-2 の対象であった東京都以外の、県議会、市議会、町議会を含む様々な規模の地方議会を対象としている点が特徴である。また、英語を対象とした Stance Classification データセットは SNS などの短い発言を対象としたものが多いが、議会議事録を対象とした本データセットは、平均 578 単語と比較的長い発言を対象とすることも特徴である。図 1 に実際の議事録の発言を示す。

愛知県の種々の規模の議会を対象に本手法を適用

したところ、6,256 サンプルの賛否ラベル付きデータを自動構築することができた。構築したデータの品質を評価するため、人手で構築したデータとの比較を行った。人手データと比べて抽出漏れが約 10% あったものの、余計な発言が抽出されることはなく、また自動ラベル付した結果に誤りは無く、高精度なデータを自動構築できることを確認した。また、構築したデータセットを基にベースライン賛否分類器を作成し、分類性能の評価を行った。

2 関連研究

スタンス分類のデータセットを構築する研究に Glandt らの研究[2]がある。この研究では、COVID-19 に関するツイートを収集し、クラウドソーシングを活用し人手でラベリングすることでデータセットを構築している。他にも、Allaway と McKeown の研究 [3]では、非常にターゲットの種類が多い VAST というデータセットを構築している。この研究でも、ラベリングにはクラウドソーシングが活用されている。クラウドソーシングの問題点として、ラベルのつけ間違いや個人差、コストがかかることが挙げられる。本研究では、明示的な賛否表現にマスクを施すことにより自動でデータセットを構築した。

3 自動データセット構築

データセットを構築する手法について述べる。まず、提出議案の討論が主に行われる会議最終日の議事録のみを用いる。地方議会の討論の進行には、議長が個々の議案を一つずつ取り上げて発言者(議員)を指名し発言を認める形式(個別討論)と、議長が発言者のみを指名した後、指名された議員が複数の議案について一括で発言を行う形式(一括討論)の 2 種類があり、議会毎に採用する討論形式が異なる。今回は発言とその対象となる議案の対応関係が明確である、個別討論形式の議会を対象とした。

愛知県の 33 の県議会、市議会、町議会を対象に調べたところ、15 の議会が条件を満たしていた。表 1 に個別議案の討論形式を持つ議会とデータセット構築に用いた議事録の対象年度を示す。

表 1 データセット構築に用いた議事録

| 対象議会 | 対象年度 |
|------|-----------------|
| 県議会 | 平成 19 年～令和元年 |
| あま市 | 平成 30 年 |
| 江南市 | 平成 15 年～令和 2 年 |
| 大府市 | 平成 2 年～令和元年 |
| 刈谷市 | 平成 3 年～平成 15 年 |
| 蒲郡市 | 平成 7 年～平成 12 年 |
| 新城市 | 平成 19 年～令和 2 年 |
| 津島市 | 平成 19 年～令和 2 年 |
| 日進市 | 平成 11 年～令和 2 年 |
| 碧南市 | 平成 19 年～令和 2 年 |
| 阿久比町 | 平成 24 年～令和 2 年 |
| 大口町 | 平成 24 年～平成 28 年 |
| 東郷町 | 平成 19 年～令和 2 年 |
| 豊山町 | 平成 21 年～令和 2 年 |
| 東浦町 | 平成 15 年～令和 2 年 |

3.1 討論の抽出

本研究で構築するスタンス分類データセットは、議事録の討論を対象とする。討論の開始・終了は議長によって指定され、討論は議長に指名された議員が行う。そのため、討論の開始表現と終了表現の間の議員の発言を討論候補として抽出する。討論候補の抽出に使用した正規表現を表 2 に示す。

スタンス分類では、特定のトピック(ターゲット)に対する意見を分類する。そのため、討論候補のうち、表 2 のトピック表現を含むものをスタンス分類のテキストとして扱い、トピック表現を含まない候補は破棄する。

得られたテキストは、冒頭や末尾付近に明確にスタンスを表明する文(賛否表明文)が現れることが多い。実際の分類タスクでは、この賛否表明文を用いるだけで簡単にスタンスが特定できてしまう。そのため、表明文は除去することが望ましいが、出現位置や表現は様々であり難しい。そこで本研究では、賛否を表明するために典型的に使用される単語集合を予め列挙しておき、それらの単語を一つの特殊なトークンに置き換える(マスクする)ことで、発言に明示的な賛否表現が含まれないようにした。今回は、「賛成」と「反対」の 2 単語を対象にマスクを行った。

表 2 本研究で使用した正規表現

| | |
|--------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 討論開始表現 | 討論(に 入り を 行い を 開始(いた)? し を 許 し ます 討論の 通告 が あ(り ま す した) る)(の で ため) |
| 討論終了表現 | ((討論(を は))?(終 わり ま 終 了(いた)? しま 省略(いた)?しま 終 結(いた)? しま)(す した))(討論を 終 結 し) |
| トピック表現 | ((議案 意見書 認定 請願)第[0-9 0- 9 一 二 三 四 五 六 七 八 九 十 百]+号 第 [0-9 0-9 一 二 三 四 五 六 七 八 九 十 百]+号議案)(「.+?」 .*?(に つ いて き ま して) に 関 し(ま し)?(て)? に 対 し(ま し)?(て)?)? |
| 賛成表現 | 賛成((の を する)立 場(で か ら))?討 論を?(さ せ て い た だ き い た し 申 し 上 げ し ます)。 |
| 反対表現 | 反対((の を する)立 場(で か ら))?討 論を?(さ せ て い た だ き い た し 申 し 上 げ し ます)。 |

3.2 賛否ラベリング

スタンス分類のデータセットには、サンプル(発言)毎にスタンス(賛成または反対)の教師ラベルを付与する必要がある。しかし、人手でラベル付けするコストは高い。本研究では、マスクした「賛成」「反対」を利用しヒューリスティクスを用いて自動でラベル付けを行った。ラベル付け手法は以下の通りである。

1. マスクした「賛成」および「反対」が発言を通してどちらか一方のみであった場合、マスクした単語が表すスタンスのラベルを付与する。
2. 「賛成」「反対」どちらも出現する場合、発言中に表2の「賛成表現」または「反対表現」が出現するかチェックする。出現する場合は、対応するラベルを付与する。
3. 上記に当てはまらない発言は、ラベル付けせずに破棄する。

表3に表1で示した議事録から構築されたデータセットの統計情報を示す。表3より、賛否ラベルの割合がほぼ等しいデータセットであるといえる。また、図2に作成したデータの例を示す。

表 3 データセットの統計情報

| | データ数 |
|----------|---------------|
| データセット全体 | 6,254 |
| 賛成ラベル | 3,437 (55.0%) |
| 反対ラベル | 2,817 (45.0%) |

テキスト：議案第38号「大府市空家等対策の推進に関する条例の制定について」、自民クラブを代表し、[STANCE]の立場で討論をさせていただきます。近年、人口減少や少子高齢化等を背景に、全国的に空き家等の増加が社会問題となっている中、本市においても、今後、空き家等の増加により、生活環境の保全に支障が生じるのではないかと懸念されているところであり、本市議会建設消防委員会においても、これまで、空き家、空き地、空き店舗について調査研究を進めてきた経緯もございます。本市では、平成30年3月に、大府市空家等対策計画を策定し、本計画に基づき、これまでも積極的に空家等対策を推進してきたところでございますが、法の枠組みだけでは対処できない課題が顕在化している状況がうかがえます。そのような状況の中で、今回、本条例を制定し、類似空家と特定類似空家等への対応を始め、市、所有者、市民、事業者などに責務又は役割を定めていること、人の生命、身体又は財産に重大な損害を及ぼすことを回避するために、緊急安全措置について規定していることは、法律を補完するとともに、空家等対策を総合的かつ計画的に推進し、公共福祉と地域振興につながるものと高く評価するものでございます。今後、法及び本条例に基づいて、引き続き空家等対策に継続的に取り組み、市民が安心安全に暮らすことのできる「健康都市おおぶ」として、持続可能な社会が実現されることを期待し、[STANCE]討論とさせていただきます。

ターゲット：議案第38号「大府市空家等対策の推進に関する条例の制定について」

ラベル：賛成

図 2 作成したデータの例

3.3 データセット構築手法の評価

構築したデータセットについて、人手で評価を行った。大府市の議事録を対象とし、議員の発言を100件抽出し、人手で討論発言かどうか分類した。討論発言である場合は、ターゲットとラベルも人手で付与し、同じ100件の発言に対して自動で構築したデータセットと比較した。比較結果を表4に示す。

表4より、自動抽出では、本来抽出すべき討論の約10%を取りこぼしていることがわかる。しかし、間違えて討論でないものを抽出していないため、データの数は少なくなるが質は高いと言える。自動で抽出した討論に対するターゲットとラベルは人手と完全に一致し、ミスなく付与できている。

表 4 人手評価の結果

| | 自動で抽出 +賛成付与 | 自動で抽出 +反対付与 | 自動で抽出 しなかった |
|----------------|----------------|----------------|----------------|
| 人手で抽出+ 賛成付与 | 41 | 0 | 4 |
| 人手で抽出+ 反対付与 | 0 | 26 | 5 |
| 人手で抽出 しなかった | 0 | 0 | 24 |

4 ベースライン分類器の構築と評価

4.1 ベースライン分類器

構築したデータセットを学習データに用いて分類器を構築した。分類モデルには、Transformer エンコーダベースの事前学習モデルである BERT[4]を用いた。BERT モデルには、東北大学乾研究室が公開している学習済みモデルⁱを使用した。[CLS]トークンを対象に、MLP 層を加えた二値分類器を学習した。入力形式は、対象となる議案名と、発言テキストを、[SEP]トークンで連携したものとした。ただし、BERT の最大入力長である 512 トークンを超えた場合は、後を切り捨てた。

4.2 分類器の評価実験

構築したデータセットから 3.3 節で人手評価した 67 件を除いた 6,187 件を学習データとして、4.1 節のモデルを 6 エポック学習し、テストデータ 67 件を評価した。分類結果を表 5 に示す。

表 5 ベースライン分類器での評価結果

| | 正答率 |
|-------|--------|
| すべて賛成 | 0.6119 |
| BERT | 0.8657 |

5 おわりに

議会議事録を活用して、スタンス分類のデータセットを自動で構築した。人手評価の結果、取りこぼしているデータは多少あるが、ラベリングやターゲットの付与は非常に高い精度であることがわかった。今後の研究としては、今回対象としなかった、複

数の議案を含む発言からもデータセットを構築することを目標とする。そのためには、発言を議案ごとにセグメンテーションする手法を研究する必要がある。

本研究で構築したデータセットは、NTCIR-17 QA Lab-PoliInfo-4[5] のサブタスク Stance Classification-2[6] のデータセットとして使用する予定である。

参考文献

1. Yasutomo Kimura, Hideyuki Shibuki, Hokuto Ootake, 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. Kyle Glandt, Sarthak Khanal, Yingjie Li, Doina Caragea, and Cornelia Caragea. 2021. Stance Detection in COVID-19 Tweets. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), pages 1596–1611, Online. Association for Computational Linguistics.
3. Emily Allaway and Kathleen McKeown. 2020. Zero-Shot Stance Detection: A Dataset and Model using Generalized Topic Representations. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 8913–8931, Online. Association for Computational Linguistics.
4. Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. 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), pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.
5. 小川泰弘, 木村泰知, 渋谷英潔, 乙武北斗, 内田ゆず, 高丸圭一, 門脇一真, 秋葉友良, 佐々木稔, 小林暁雄. NTCIR-17 QA Lab-PoliInfo-4 のタスク設計. 言語処理学会第 29 回年次大会. 2023.
6. 高丸圭一, 内田ゆず, 木村泰知, 秋葉友良. 地方議会における議案への賛否に関する発言の分析—

ⁱ <https://huggingface.co/cl-tohoku>

NTCIR-17 QA Lab-PoliInfo-4 Stance Classification-2
タスクに向けて一. 言語処理学会第 29 回年次大会.
2023