

地方議会特有の表現に基づく政策に対する発言の賛否分類

勝山 光[†] 杉本 翔[‡] 田上 諒[‡] 木村 輔[‡] 宮森 恒[‡]
 京都産業大学 コンピュータ理工学部[†] 京都産業大学大学院 先端情報学研究科[‡]

{g1544377, i1888097, i1788124, i1658047, miya}@cc.kyoto-su.ac.jp

1 はじめに

近年、フェイクニュースの影響により、既出の情報について真偽を検証する、**事実検証**の重要性が再確認されている。例えば、フェイクニュースのターゲットとなりやすい、政治家の発言について事実検証するとき、地方議会議録のような、**一次情報源**を確認する必要がある。しかし、このような会議録は、議事の進行や提出された資料を含む、膨大な文量を有しているため、議員の意見を一目で確認することは難しい。

国立情報学研究所が主催するワークショップのNTCIR-14では、近年のフェイクニュースの問題に対するタスクである、QA Lab-PoliInfo Taskのサブタスクの1つとして、ある政策についての地方議会中の発言が、**根拠を伴う賛成意見**、**根拠を伴う反対意見**、および、**それ他の**どれであるか分類するClassification Taskを実施している。具体的には、図1に示すように、**会議録中の発言文 (Utterance)**、および、**政策の課題文 (Topic)** から、**関連性**、**事実検証可能性**、および、**意見性**の3種類のラベルについてそれぞれ分類し、**Topic**に対する**Utterance**の最終的な**賛否**を判断する。なお、簡単のため、根拠を伴う賛成意見を単に**賛成**、また、根拠を伴う反対意見を**反対**と呼ぶこととする。

本稿では、QA Lab-PoliInfoにおけるClassification Taskを対象とした、**地方議会特有の表現から生成した分散表現および素性**に基づく、各ラベルの分類手法をまとめる。

2 関連研究

2016年に、フェイクニュースに関わる問題の解決を目指し、世界中の学会、および、業界から、100人を超えるボランティアと71のチームが協力し、Fake News Challenge (FNC) という取り組みが誕生した。その最初の取り組みとして、Fake News Challenge Stage

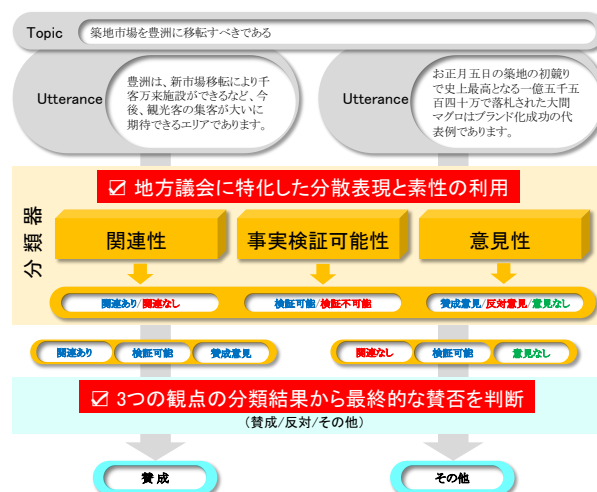


図 1: 提案手法の概要

1 (FNC-I) が開催された。この取り組みでは、ある記事と見出し文のペアが入力として与えられ、その見出し文が、記事の内容に対して、合致している、合致していない、議論している、関係がない、の4つの立場について分類を出力するタスクが設計されている。

このタスクにおいて、全80チームのうち、最高精度を収めたSeanら[3]は、まず、畳み込みニューラルネットワークによる分類モデルと、弱分類器に決定木を用いた勾配ブースティングによる分類モデルを構築した。そして、各モデルの出力である確率値を、1対1の加重平均で組み合わせ、最終的な分類を決定する手法を提案している。それぞれのモデルを単一で推定した場合と比べ、両方のモデルを組み合わせることによって、精度が向上したことを報告している。また、2位のモデルを提案したHanselowskiら[1]の手法では、uni-gramの出現頻度に基づくBoWベクトルをTFIDF値に置き換えたベクトルに、複数の手法で獲得したトピックベクトルなどを連結したベクトルを入力とする、多層パーセプトロンの分類モデルを5つ学習し、それぞれの推論結果の投票により分類する手法を提案している。

3 提案手法

ここでは、提案手法である**地方議会特有の表現から生成した分散表現および素性**に基づく、地方議会における発言の賛否分類手法について詳述する。本分類手法は、図1に示すとおり、入力として、**Utterance**、および、**Topic**が与えられたとき、まず、**関連性**、**事実検証可能性**、および、**意見性**の3つの観点についてそれぞれ分類する。その後、各分類結果から、表1に示すルールに従って、**Topic**に対する**Utterance**の最終的な**賛否**の判断を出力する構成となっている。

表 1: 3つの観点の分類結果の組み合わせによる賛否分類

賛否	関連性	事実検証可能性	意見性
賛成	関連あり	検証可能	賛成意見
反対	関連あり	検証可能	反対意見
その他	(それ以外の組み合わせ)		

3.1 関連性

ここでは、ある**Topic**に対する**Utterance**について、その発言内容が**Topic**に、**関連する**、または、**関連しない**を判断する、2値分類器を構築する。

本研究では、分類器として、Joulinら[2]の手法を適用した。この手法は、分かち書きされた一文を入力とし、確率値を出力する、一層から構成されるニューラルネットワークである。我々は、**Topic**と**Utterance**の文を連結した一文を入力とした。

3.2 事実検証可能性

ここでは、ある**Utterance**について、その発言内容が、**事実確認できる**、または、**事実確認できない**を判断する、2値分類器を構築する。

本研究では、文中に、場所や日付、金額などの表現が含まれ、事実の検証が可能である**Utterance**を、**事実確認できる**と定義する。分類器は、深層学習によるモデルとし、**Utterance**をLong-Short Term Memory (LSTM)でエンコードした上で、全結合層によって、2値の確率値を出力する構造とする。

3.3 意見性

ここでは、ある**Topic**に対する**Utterance**について、その発言内容が、**賛成意見**、**反対意見**、および、

意見なしのうち、どの意見を表すのかを判断する、3値分類器を構築する。

提案手法では、まず、賛成意見と反対意見をまとめた**意見あり**、および、**意見なし**について判断する2値分類器を構築し、その後、意見ありと判断された発言のみを対象に、**賛成意見**、および、**反対意見**を判断する2値分類器を構築した。予備実験を踏まえ、分類器には、Support Vector Machine (SVM)を選択した。

分類に用いた各素性は、大きく2種類あり、単語単位のN-gram (N = 1,2,3)の出現頻度に基づく素性3タイプ、および、係り受けに基づく文節単位の2-gramの出現頻度に基づく素性3タイプである。ここで、文節単位の2-gram素性は、文節として扱う単語列について、内容語(名詞、形容詞、動詞) + 機能語を文節と扱う**条件1**、内容語のみを文節と扱う**条件2**、動詞のみを文節と扱う**条件3**の3タイプを作成した。

各素性について、それぞれの2値ラベルにおける出現頻度、上位200件、400件、600件を選択し、素性の次元数とした。例えば、賛成意見/反対意見分類の場合、賛成意見ラベルが付与された全てのデータにおける出現頻度上位100件、および、反対意見ラベルが付与された全てのデータにおける出現頻度上位100件を合わせた、計200件を素性の次元数とした。

また、次元数200の各素性について組み合わせ、以下の素性を構築し、計31種類の素性を構築した。まず、単語単位のN-gramの素性の2種類を組み合わせ、次元数400の素性を4種類作成した。さらに、N = 1,2,3のそれぞれの素性を組み合わせ、次元数600の素性を1種類構築した。最後に、文節単位の2-gramの素性と、単語単位のN-gramの素性から、それぞれ1つを選択し組み合わせ、次元数400の素性を9種類作成した。

3.4 データセット

ここでは、各分類器を構築する際に用いた、QA Lab-PoliInfo Classification Taskの地方会議録における、各政策への賛否意見データセットについて詳述する。このデータセットは、14種類の**Topic**と、各**Topic**のキーワードを含む**Utterance**のペア10,289件に対して、関連性、事実検証可能性、意見性、及び、それら3ラベルから判断される賛否ラベルが付与されている。

各ペアのそれぞれのラベルは、3人、または、5人の評価者より付与されているため、本研究では、次の正解ラベル決定方法を用いて、開発データにおける正解

ラベルを付与した。**タイプ1**では、複数人から付与された各ラベルについて、多数派のラベルを正解ラベルとする。また、3.1節の**関連性**については、次の正解ラベル決定方法によって構築されたデータセットについても分類器を学習した。**タイプ2**では、複数人から付与された各ラベルについて、一人でも「関連なし」を付与していたら、正解ラベルを「関連なし」とする。

4 実験・考察

提案手法における最終的な賛否の判断は、**関連性**、**事実検証可能性**、および、**意見性**の各ラベルの組み合わせによって決定される。そのため、3.1節～3.3節のそれぞれの分類器の精度について確認する必要がある。本実験の目的は、より適切なモデルの組み合わせを決定するために、それぞれ提案した分類手法について、その精度を調査することである。

各分類器における、入力データ (**Topic**、**Utterance**) は、形態素解析器である MeCab を用いて分かち書きした。また、3.1節の**関連性**のニューラルネットワークとして、fastText を用いた。3.2節の**事実検証可能性**の分類器は、Chainer によって実装し、3.3節の**意見性**の係り受け解析器として CaboCha を用い、機械学習ツールとして Weka を用いた。

4.1 関連性

実験のためのデータセットのラベル決定方法として、**タイプ1**、及び、**タイプ2**を用いた。また、評価指標として適合率、および、再現率を用い、10分割交差確認法により評価した。fastText のデフォルトパラメータのうち、予備実験より、単語 N-gram、単語の特徴次元数、エポック数、学習率、及び、誤差関数を、それぞれ、2-gram、50次元、50エポック、1.0、階層的ソフトマックスに決定した。

以上の条件における評価結果を、表2、および、表3に示す。異なるラベル付与基準のデータセットか

表2: **タイプ1**における関連性を判断する分類モデルの10分割交差確認法の結果

ラベル	適合率	再現率	F 値
関連あり	0.950 (9,147/9,624)	0.974 (9,147/9,388)	0.962
関連なし	0.638 (424/665)	0.471 (424/901)	0.542

表3: **タイプ2**における関連性を判断する分類モデルの10分割交差確認法の結果

ラベル	適合率	再現率	F 値
関連あり	0.912 (676/7,289)	0.933 (676/7,121)	0.922
関連なし	0.842 (175/3,000)	0.797 (175/3,168)	0.819

ら学習しているため、一概に比較することはできないが、表2、および、表3より、開発データの不均衡を軽減した**タイプ2**について、関連なしに対する適合率、および、再現率が大きく改善していることが確認できる。一方、**タイプ1**については、関連ありに対する各評価指標は高いものの、関連なしの評価指標と大きく差が開いた。分類結果を確認したところ、関連なしラベルの数が2桁以下の政策10件中8件については、すべての発言を関連ありと判断していることが確認できた。

4.2 事実検証可能性

実験には、関連性と同様に、3.4節のデータセットを用い、ラベルの決定方法は、**タイプ1**のみを使用した。LSTMによるエンコード時には、我々が独自に収集した東京都議会本会議全1,209件をもとに生成した、単語の分散表現を用いた。パラメータとして、エンコード後の中間ベクトルは50次元、全結合層は1層で、活性化関数はソフトマックス、エポック数は20、誤差関数は交差エントロピーを選択した。最適化アルゴリズムに Adam を使用し、各パラメータを $\alpha = 0.001$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 10^{-8}$ とした。データセットのうち、9割を学習データ、1割を検証データとした。

以上の条件で学習させたところ、最終エポック終了時点で、検証データに対する Accuracy は、約82.7%となった。詳細な結果を表4に示す。表4より、学習されたモデルは、分類結果として、事実確認できないと判断しやすいことが分かる。これは、学習データ中のラベルの割合が不均衡であるためだと思われる。ただし、検証データにおいても、割合はほぼ同一のため、正答率の大幅な低下にはつながっていない。

4.3 意見性

実験のためのデータセットのラベル決定方法として、**タイプ1**を用い、また、評価指標として、適合率、再現率、および、F 値を用い、10分割交差確認法により

表 4: 事実検証可能性を判断する学習済みモデルを用いた検証データに対する分類精度

ラベル	適合率	再現率	F 値
確認できない	0.847 (676/798)	0.923 (676/732)	0.883
確認できる	0.757 (175/231)	0.589 (175/297)	0.662

評価した。意見あり/意見なし、および、賛成意見/反対意見の分類結果について、各素性の特徴を比較するために、3.3 節で示した各素性のうち、次元数 600 の 6 モデルを比較することとした。

各分類器の実験結果を、表 5、および、表 6 に示す。いずれの分類器においても、uni-gram を素性とした

表 5: T:意見なし/F:意見ありを判断する分類モデルの 10 分割交差確認法の結果

素性	適合率		再現率		F 値
	T	F	T	F	
uni-gram	0.904	0.664	0.963	0.418	0.870
bi-gram	0.888	0.702	0.978	0.294	0.854
tri-gram	0.878	0.762	0.988	0.215	0.841
条件 1	0.884	0.744	0.984	0.258	0.850
条件 2	0.885	0.720	0.981	0.274	0.851
条件 3	0.860	0.597	0.991	0.074	0.803

表 6: T:賛成意見/F:反対意見を判断する分類モデルの 10 分割交差確認法の結果

素性	適合率		再現率		F 値
	T	F	T	F	
uni-gram	0.636	0.941	0.383	0.978	0.916
bi-gram	0.666	0.989	0.223	0.957	0.901
tri-gram	0.625	0.927	0.223	0.987	0.899
条件 1	0.686	0.927	0.220	0.990	0.901
条件 2	0.659	0.929	0.234	0.988	0.902
条件 3	0.622	0.914	0.060	0.996	0.877

モデルが、F 値において最も高い精度となった。特に、他のモデルと比較して、意見あり、および、賛成意見の再現率が高いことがわかる。一方、bi-gram の素性や条件 2 の素性などの、2 単語、または、2 文節以上の関係を捉えた素性を用いたモデルの精度は、uni-gram 素性の精度を大きく改善することは確認できなかった。各素性において用いた出現頻度上位の語句を調べたところ、uni-gram の素性においては、主語である「私」や、賛否の立場を表現する「反対」や「求める」といった語句が多く存在することが確認できた。一方、それ以外の素性においては、例えば、「反対、です」や「反対、する」など、もともと単一だった特徴的な表現に多様性が生じていることがわかった。以上の原因により、

素性中に含まれる特徴的な語句の割合が低下したため、uni-gram 素性に劣る結果になったと考えられる。

5 まとめ

本論文では、フェイクニュースなどの事実検証手法として、地方議会会議録中のある政策に対する発言を対象とした立場分類手法を提案した。提案手法では、政策と発言間の**関連性**、発言内容の**事実検証可能性**、政策に対する発言者の**意見性**について、**地方議会特有の表現から生成した分散表現および素性**に基づく、それぞれ分類器を構築し、各分類器の実験結果をまとめた。

謝辞

本研究の一部は科研費 18K11557 の助成を受けたものです。ここに記して感謝の意を表します。

参考文献

- [1] Andreas Hanselowski, Avinesh PVS, Benjamin Schiller, and Felix Caspelherr. Description of the system developed by team athene in the fnc-1. https://github.com/hanselowski/athene_system/blob/master/system_description_athene.pdf, 2017. Accessed: 2018-01-16.
- [2] Armand Joulin, Edouard Grave, Piotr Bojanowski, and Tomas Mikolov. Bag of tricks for efficient text classification. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, pp. 427–431. Association for Computational Linguistics, 2017.
- [3] Baird Sean, Sibley Doug, and Pan Yuxi. Talos targets disinformation with fake news challenge victory. <http://blog.talosintelligence.com/2017/06/talos-fake-news-challenge.html>, 2017. Accessed: 2018-01-16.