

地方議会会議録における要求・要望表現抽出の提案

葦原 史敏[†] 木村 泰知[†] 荒木 健治[‡]
[†]小樽商科大学 [‡]北海道大学大学院情報科学研究科

1 はじめに

近年、自治体における住民参加（パブリックインボルブメント）の重要性が認識されてきている[1]。そのような中、地方自治の根幹をなす地方議会にも住民参加を推進する役割が期待されている。しかしながら、地方議会に対する住民の評価は低く、その理由は「議会が何をやっているかわからない。」というものが最も多い[2]。そのため、地方議会に関わる情報をわかりやすく住民に提供することは、住民参加を推進する上で重要と考えられる。

これに対し、我々はこれまで先行研究で行われてきた要求・要望表現の抽出[1][3]を地方議会会議録（以下、会議録）に対して適用することで、住民にとって重要である、地方自治の問題を獲得できると考えた。

特に会議録中では、表現が話し言葉で、一文が非常に長いものが存在するという特徴がある。このような場合、一文に複数の事態が記述されており、先行研究[1]で行われてきたような一文単位での要求・要望表現抽出ではその文中に存在する個々の事態に対応できない。そこで、本研究ではこの問題に対応するために節単位での要求・要望表現判定を行うことを提案する。この際、先行研究のように注釈者間の一致が保障された注釈付け、精度の良い要求・要望表現の推定が可能か、という点が問題であり、本稿ではそれらについて検証していく。なお、地方自治の問題を獲得するという目的のために、対象とするデータは会議録の中でも議員の質問に限定する。

以下、2では会議録中における要求表現の分析を行い、人手による要求・要望表現判定の一致調査の結果を示す。3では節分類の情報などを素性とした機械学

習による要求・要望表現判定の実験結果を示す。最後に4で本稿の内容をまとめる。

2 要求・要望表現の注釈付け

本節では、要求・要望表現を注釈付けする際に問題のある会議録中の表現の例を示し、その表現の例に対応した要求・要望表現の注釈付けの枠組みを提案する。

次に、この枠組みに基づいて注釈付けを行った際の一致率を評価し、適切な注釈付けが可能か考察する。

2.1 会議録の分析

会議録には一文単位による要求・要望表現の注釈付けでは対応できない例が多く存在する。また、地方自治で問題となっている事柄を獲得するという目的にそぐわない要求・要望表現が存在する。以下にその例を示す。

一文が非常に長い例

会議録中には(1)の例のような非常に長い一文が存在する。このような文から要求・要望表現を抽出する場合、文末に着目するだけでは、下線に示すような内部に存在する要望を表す表現を対象にすることができない。

- (1) 介護保険制度でぜひ改善してほしいことについては、診療報酬を改定してほしい、特に、ヘルパーの事業所では収支が合わない、2級ヘルパー、介護福祉士も、一区切りになっているために、同料金の収入ですけれども、給料には差をつけるしかない、また、冬期の事業所では、雪のある地域では除排雪費がどうしても出ていく、そういう地域の人件費のそういうものも検討してほ

しいという意見が寄せられました。

説明の要求の例

会議録中には、(2)のような説明を求める表現がある。しかし、本研究の目的である地方自治における問題の獲得にそぐわないため、カテゴリを分ける必要がある。

(2) そこで、質問ですが、小樽市の商店街の現状認識と、将来展望に向けたあるべき姿をどのようにとらえているのか、お示してください。

説明の要求における婉曲的な表現の例

先行研究[1]において「してほしい」という表現で言い換えられるものは明示的な要求と定義されている。しかし、(3)のような婉曲丁寧な表現は言い換えると不自然になるため、先行研究のように明示的な要求のみを扱う場合、対応できない。

(3) 教育現場がもたらす、その社会的重要性、重大性を含め、教育長の御所見をお伺いいたします。

2.2 要求・要望表現の分類と定義

これまで述べてきた事例に対応するために、抽出単位を述語が構成する節を範囲とし、要求・要望表現を3つに分類する。

まずカテゴリは説明の要求に対応するために以下のように要求・要望表現の分類を行った。

1. 明示的説明要求
2. 暗黙的説明要求
3. 明示的問題解決要求・要望

本研究では、上記の1及び3の明示的な要求・要望を次のように定義する。

「述語が要求・要望する相手の行動を示し、その行動を話者が望んでいるモダリティを持つ」

この定義は、第一著者が会議録中で自治体・その他の行動を求める要求・要望の事例を見て妥当と判断し

た基準である。この基準に基づいて注釈付けをし、先行研究[1]のような言い換えによる判定が必要か考察する。

また、1,2は議員の質疑において、相手(市長・役人)に説明を求める表現である。

図1はこれらのカテゴリの例である。

次に、介護保険制度についてです。 . . . 小樽市として、 <u>・・現行の介護保険制度で要支援者を含むすべての要介護者へのサービスを続けるように求めます</u> 。 (明示的問題解決要求) 市長の見解をお聞かせください。 (明示的説明要求) . . . 市長の御見解をお伺いいたします。 (暗黙的説明要求)

図1 要求・要望表現の分類例

次に注釈単位であるが、範囲を規定するために述語を表1のものに指定する。なお、3列目の出現頻度は述語数1000における頻度内訳である。

表1. 述語の定義とその出現頻度

述語	例	出現頻度
動詞	つくる, 示す, 聞くなど	840
です, ある	「～の質問です」 「～であります」など	108
ください・いたす	「お答えください」 「質問いたします」など	52

これらの述語を含む節を注釈付けの範囲とする。(4)に例を示す。

(4) 厚生労働省から福祉避難所設置・運営に関するガイドラインも出されており、小樽市でも福祉避難所をつくる必要があります。

注釈者は述語に着目し、その述語の表現する事態を

発言者が望んでいるか、という基準で要求・要望か判定する。具体的には「つくる」という述語に対し、「必要がある」というモダリティ表現が付与されていることから要求・要望表現であると判定する。この際、省略語などの文脈情報は前後から確認して良い、としている。

2.3 注釈付けの一致率評価

以上の規定を踏まえて、文系大学生3人(男性1人、女性2人)と第一著者の間で注釈付けの一致率を計る実験を行った。データは平成23年度小樽市第2回定例会の議員の質問からランダムに述語を200選択したものをを用いる。ただし、要求・要望表現は述語全体における割合が低いため、その内一部(十数個)を第一著者が要求・要望表現と判定したものに置き換えた。一致率はコーエンの一致係数を用いて計算する。表2にその結果を示す。

表2. 著者と注釈者I, II, IIIとのコーエンの一致係数

注釈者	コーエンの一致係数
著者と注釈者 I	0.618
著者と注釈者 II	0.376
著者と注釈者 III	0.362

一致係数の値を先行研究[1]と比較した場合、若干低くなっているが、ある程度の一致を確保できており、節の範囲を対象とした場合も注釈付けは可能であると判断できる。

2.4 不一致箇所の分析

一致していなかった箇所には、これから質問する内容を宣言する意味の文に暗黙的説明要求のタグが付与されていたものが5つあった。

(5) 最後に、港湾問題について伺います。

(5)は発語時点で具体的な説明を求めているため、説明の要求にはあたらない。これに対しては、前後を確認する「次に・最後に」といった副詞をもとに宣言的な意味かを判断することができるため、注釈付けの

一致率を改善することができると考えられる。

また、明示的問題解決要求とタグなしで判断がわかれた例も多くあった。特徴的な例は、(6)のようなものである。

(6) 一人一人の行動障害の特性を踏まえ支援しなければならず、突発的に起こるリスクに対しても的確に判断し、対応しなければなりません。

この文の場合、自治体に対して行動を求めているのか、一般論的に自治体が行うべき行動を述べているのかが理解しにくい。このような例に対しては言い換えによる明確な基準などの導入を検討する必要がある。

3 要求・要望表現の推定

3.1 実験設定

注釈付けしたデータをもとに機械学習による推定を行う。推定は要求タグ(全てのカテゴリを含む)か非要求タグかを判別する2値分類問題とする。対象とするデータは平成23年度小樽市第1回定例会の述語2000とし、10分割交差検定により精度を検証する。機械学習は分類性能の高いサポートベクタマシン¹を採用し、素性は次の3つを用いた。

[1] 述語を含む節のBOW(unigram)と次の節の品詞情報

節の範囲はCBAP[4]を用いて付与する。また、節間の接続部分の構文情報を取り入れるために、次の節の品詞情報として、述語を含む節の最後尾にある単語と、次の節の先頭2単語の計3単語に対する品詞情報のtrigramを用いる。

[2] 該当する節の分類

特定の節に要求・要望表現が現れない傾向を取り入れるために、CBAPの出力する節の分類を以下の9種類のカテゴリとして入力する。

並列節、条件節、譲歩節、理由節、時間節、その他連用節、補足節、連体節、その他

¹ ライブラリはLIBSVMを用いた

² 品詞情報はMecab[5]の品詞細分類を用いる

[3] 述語の主格

述語の主格を3カテゴリに分類して付与する。なお、データは人手で入力している。

A. 一人称

「私」「我々」など話し手である議員自身を指す主格

B. 自治体に関係する主体

自治体、役所など自治体自身を指すもの、または市長、役人など聞き手を指すもの

C. それ以外の主体

[2]と[3]の素性は該当カテゴリが出現した場合1、出現していないなら0として入力する。

また、サポートベクタマシンのカーネルは、実験の結果、精度の良かった3次の多項式カーネルに設定した。

3.2 実験結果

表3に実験結果を示す。

表3. 素性の組み合わせと精度

素性の組み合わせ	適合率 (%)	再現率 (%)	F 値 (%)
[1]	91.22	52.19	65.07
[1]と[2]	88.35	56.60	68.47
[1]と[2]と[3]	91.78	60.43	71.89

3.3 実験結果の考察

実験結果は総じて再現率が低いという結果となっている。カテゴリ別に観察すると、説明要求は表現がある程度固まっているため判別に成功しているが、一方で問題解決要求の多くは判別に成功していない。これは頻出ししない表現が多く、またモダリティ情報（「～が必要である」など）の取り入れを適切に出来ていないことに起因する。モダリティ情報は連節している際に述語の直後に出現しない場合や、出現したとしても間接疑問で要求・要望表現に当たらないものがあり、その取り入れは困難である。

素性については、節分類と述語の主格の情報を与えることで、F値において6.82%の向上が見られた。これ

は会議録中の要求・要望表現における述語の主格は自治体や市長など政治主体に偏っているためと考えられる。

4 おわりに

本稿では、会議録を対象にした要求・要望表現の抽出手法の提案を行った。特に、節の範囲で注釈付けを行う際の問題点について検証した。

注釈付けに関してはコーヘンの一致係数で0.3~0.6の値が確認され、節の単位で適切な注釈付けが可能であることが確認された。しかし、いくつか見られた注釈付け上の問題点については今後ルールを整備していく必要があると考えられる。また、正解データの正例と負例の比率が約1:10と偏っているため、対象とする述語を減らすことも検討する必要がある。

要求・要望表現の推定については、モダリティ情報の取り入れが難しく、再現率が低くなることが確認された。今後この問題点について、さらに素性の与え方や構文情報の取り入れなどを検討していく予定である。

謝辞

本研究の一部は科研費 22300086 の助成を受けたものである。

参考文献

- [1] 大塚裕子, 内山将夫, 井佐原均. 自由回答アンケートにおける要求意図判定基準, 自然言語処理 Vol. 11, No. 2, pp21~pp61, 2004.
- [2] 江藤俊昭, 地方議会制度改革の現状と今後の課題—第29次地方制度調査会答申を中心に— 国際文化研修 2010 冬 vol. 66 pp37~pp43, 2010.
- [3] 金山博, 那須川哲哉. 要望の対象の同定, 言語処理学会第14回年次大会, pp356~359, 2008.
- [4] 丸山岳彦, 柏岡秀紀, 熊野正, 田中英輝 「節境界自動検出ルールの作成と評価」, 言語処理学会 第9回年次大会, pp517~pp520, 2003.
- [5] MeCab: Yet Another Part-of-Speech and Morphological Analyzer
<http://mecab.sourceforge.net/>