

地方議会会議録の解析による社会課題の可視化

池上 十五[†] 苅米志帆乃^{††}

^{††} 長野工業高等専門学校電気電子工学科 〒381-8550 長野県長野市徳間 716

E-mail: [†]15203@g.nagano-nct.ac.jp, ^{††}s_karikome@nagano-nct.ac.jp

あらまし 現在、各都道府県議会や市町村議会などでの地方自治体の議会では各地方で問題とされている社会課題について日々議論されている。会議録は web 上で公開されており、これを閲覧することで、住民は「社会課題」を知り、選挙などで参考にすることができる。しかし、これらの会議録は実際の議会での発言を文字起こしして作られているため、口語表現で記されており、一見して議論されている「社会課題」を判断することが困難である。そこで、会議録から「社会課題」を自動で抽出し、可視化を行う手法を提案する。

キーワード 地方議会、会議録、社会課題、可視化

1 はじめに

地方議会ではその自治体の首長や議会に所属する議員らによって「社会課題」について日々議論されている。この議会の会議録は各都道府県・市町村がそれぞれ web 上で公開している。議会では一般に特定の「社会課題」について首長や委員会の委員長、担当部署の責任者らが報告・提案を行ったり、それらの人物に対して所属議員が質問・提案し、それに首長側が回答するといった内容が主である。会議録はこの内容がそのまま文字起こしされている。

会議録は各自治体で長期にわたって保管・公開されている。これを閲覧すればその地域の住民は議会で議論されている「社会課題」について知ることができる。しかし、先述のように会議録は議会での発言がそのまま文字起こしされたデータであるため、これを読むだけでは一見して議論されている「社会課題」を見つけることは困難であり、時間がかかる。

そこで、本研究では会議録で議論されている「社会課題」とその流れを簡単に理解できるように、会議録を解析し、そのなかから「社会課題」を可視化する手法を提案する。また、情報の補足のためにそれらを「教育」「農業」「文化・福祉」といったカテゴリを付与する。可視化のイメージを図 1 に示す。図のように、カテゴリを選択するとそのカテゴリに分類された社会課題の一覧が表示される。

先行研究として、木村ら [1] は議会会議録について特徴の整理やどのように要約するかを検討した。また、乙武ら [2] は述語項構造解析を利用し会議録中から理由付き意見を自動抽出する手法を提案している。さらに、高丸ら [3] によって、会議録から発言議員名・発言の内容を取得し、TF・IDF 法によって議員ごとの特徴語を取得、データベース化する研究がされている。

上記の研究は社会課題を抽出することが目的ではないが、本研究と同様に政治問題の抽出を目的として木村ら [4] は、政治問題において重要な役割を持つ形態素である「主辞」を登録することで、それに基づいた抽出を行う方法を提案している。しかし、この研究は修飾語を含むような長い政治問題を対象とし

た研究であり、一見して「対象」とその「目的」を理解したい場合には適さない。

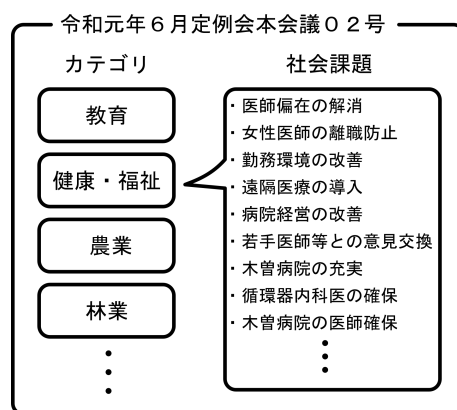


図 1 可視化のイメージ

2 「社会課題」の定義

2.1 「社会課題」の定義

本研究では、下記の 2 点を満たすものを「社会課題」と定義する。

- (1) 自治体における施政や政策決定の上で考慮が必要な問題・懸案事項
- (2) 対象とその問題が明確である

2.2 抽出パターン

会議録中において前節の定義を満たすパターンは数多く存在する。それらの中には修飾語を多く含むような長い社会課題が含まれており、これらは一見して趣旨を理解することが難しい。また、抽出された社会課題の粒度が揃っていないと一覧性が低く、本研究の目的にそぐわない。よって、本研究では抽出するパターンを定める。抽出するパターンは以下の 2 パターンである。

- (1) 「名詞+の+名詞」「名詞+への+名詞」など、名詞間をつなぐ助詞に「の」を含むパターン

(2) 「名詞+による+名詞」となるパターン

なお、上記の名詞は接頭詞のついた名詞、複合名詞を含む。また、これらのパターンと合致するものを社会課題の候補とし、そのうち前節の定義を満たすものを社会課題とする。

3 会議録の解析手法

3.1 概要

社会課題を可視化するため、会議録を解析し社会課題の抽出とカテゴリ付与を行う。手法のフローを図2に示す。まず、2.2節のパターンに従って社会課題の候補を抽出する。つづいて候補の中から2.1節の定義を満たさず、社会課題として認められないものを除外する。残った社会課題に対してカテゴリを付与し、可視化を行う。

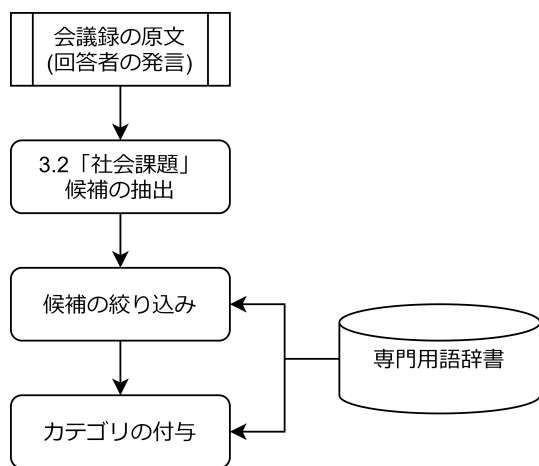


図2 可視化のためのフロー図

3.2 抽出範囲

議会はある社会課題について質問・提案をする委員会所属の議員と、それに対して回答を行う知事や委員長など首長側の人物が、交互に発言する形で進行されることが一般的である。ここでは前者を「質問者」、後者を「回答者」と呼称する。質問者と回答者が言及している社会課題は同一であるため、本研究では回答者から抽出を行う。

会議録のフォーマットは各自治体で異なるが、誰がどの発言をしたかを明確にするために、発言の前には発言者名とその役職が記されている場合が一般的である。また、発言者が回答者や質問者、議長などのうち、どのような立場であるかを示すために、ほとんどの会議録では発言者名の前に記号が付されている。下記に長野県議会会議録[5]の一部を示す。

○議長（清沢英男 君）

次に、行政事務一般に関する質問及び知事提出議案を...

◆34番（依田明善 君）

おはようございます。自由民主党県議団、南佐久郡選出の依田明善でございます...

きょうの私の一般質問は、農作業に従事する日本人がなぜ激減してしまったのか、なぜ外国人に頼るようになってしまったのか...

◎教育長（原山隆一 君）

五感を使う労働の大切さについてのお尋ねでございます。議員御指摘のとおり、子供たちが教室を...

この例の場合、長野県議会の場合は「○」が議長、「◆」が質問者、「◎」が回答者である。このような記号によって発言者の立場が特定できる。本手法ではこの記号を用いて、回答者記号が出現した後から次の記号が出現する部分までを回答者の発言として判断し、この部分から抽出を行う。

3.3 社会課題候補の抽出

まず、社会課題の候補を抽出する。2.2節より、抽出するパターンは名詞・接頭詞・助詞によって構成される。形態素とその品詞情報を把握するために形態素解析を用いる。まず、形態素解析を行う。解析結果より接頭詞・名詞を見つけ、これらが連続する場合は複合名詞として判断する。この名詞が出現した後に、「の」が含まれる助詞の連続、「による」が出現し、且つそのあとに続く形態素が名詞・複合名詞の場合、それらをひとつの候補として抽出する。

3.4 社会課題の絞り込み

抽出した候補は、回答者の発言から2.2節の形式を抜き出したただけで、2.1節の条件を満たしていない。これらの条件を満たした社会課題を取得するために以下の手法で絞り込みを行う。

(a) 専門用語を用いた絞り込み

抽出した候補のうち、社会課題として残すべきものは対象のカテゴリで頻出する語、すなわち専門用語が含まれている場合が多い。また、専門用語の位置について

- 前方・後方の名詞の両方
- 前方・後方の名詞のどちらか

である社会課題がそれぞれ存在する。そこで、専門用語辞書を作成し、それをもとに、専門用語がどこに含まれているかを判定して絞り込みを行う。

本研究で必要な専門用語辞書の作成手順を示す。対象の自治体の過去の会議録について、名詞（複合名詞を含む）へのカテゴリ付与を行い、それぞれの名詞のカテゴリごとの出現数を算出する。これらの名詞について、

$$\frac{\text{対象のカテゴリでの出現回数}}{\text{対象のカテゴリ以外での出現回数}} \geq n \quad (1)$$

となる語をそのカテゴリでの専門用語とする。ここで、 n の値は任意とし、この値によって専門性の強弱を調整する。カテゴリ付与の方法は次節で述べる。

(b) 構成する名詞数をもとにした絞り込み

専門用語を用いた方法と同様に、残したい社会課題について「前方・後方の名詞ともに単一名詞」である場合より、「どちらかが複合名詞」である場合が多い。そこで、候補に含まれる名詞を構成する形態素数について条件を設定することで絞り込みを行う。

3.5 カテゴリの付与

社会課題、及び前節の専門用語辞書作成に用いるための名詞へのカテゴリ付与の方法を説明する。

3.2 節より、抽出された候補、及び専門用語辞書に登録する名詞は全て回答者の発言から得たものである。また、前述のとおり、一般に会議録の発言の前には回答者情報が記されている。この回答者情報には「知事」や「副知事」、「教育長」や「健康福祉部長」といった役職が含まれている。この役職を利用してそれぞれの候補にカテゴリを付与する。

事前に自治体における役職とそのカテゴリの対応を人手でリスト化する。このリストをもとに役職からカテゴリを特定し、候補に対し付与する。表 1 に長野県議会におけるカテゴリとそれに対応する役職の例を示す。

表 1 カテゴリと対応する役職の例	
カテゴリ	役職の例
教育	教育長，教育委員会委員長
健康・福祉	健康福祉部長，衛生部長
農業	農政部長
林業	林務部長
環境	環境部長
観光	観光部長
インフラ	建設部長，土木部長
警察	警察本部長
広報	企画部長，企画振興部長
文化	県民文化部長
災害	危機管理部長
企業・労働	公営企業管理者，商工労働部長
行政	人事委員会委員長，監査委員長

しかし、この手法では回答者が「知事」や「副知事」など、様々なカテゴリについて発言する役職である場合、役職をもとにカテゴリを特定することができない。このような場合、発言中に含まれる専門用語の語数をもとに、カテゴリ特定を行う。発言中の語それぞれについて前節の専門用語辞書を参照し、専門用語が最も多かったカテゴリをその発言中での社会課題に対して付与する。

4 予備実験

4.1 概要

3.4 節の絞り込み手法において、各パラメータ・条件設定を決定するために、予備実験を行った。対象は長野県議会会議録「令和元年 6 月定例会本会議 02 号」の一部である。文字数は 20,282 文字であった。この中から社会課題の定義に従って、抽

出したい社会課題を人手で選定した結果、46 件が含まれていた。これを正解データとする。

4.2 方法及び結果

まず、辞書作成を行った。元データには平成 18 年 2 月～令和元年 6 月の長野県議会定例会の会議録計 458 件を使用した。この中に対象となる（専門用語の候補となる）語は 2,000,160 語含まれている。ここで、(1) 式は $n = 1$ 、 $n = 1.2$ としてそれぞれ作成した。

次に、作成した辞書を用いて専門用語に関する制限・名詞数に関する制限それぞれの条件を変えて適用し、抽出した。また、その際の精度・及び再現率を調査した。ここに、精度及び再現率は以下の式を用いて算出した。

精度 =
$$\frac{\text{自動抽出した社会課題のうち正解データと一致した数}}{\text{自動抽出した社会課題の数}} \quad (2)$$

再現率 =
$$\frac{\text{自動抽出した社会課題のうち正解データと一致した数}}{\text{正解データの数}} \quad (3)$$

表 2、3 に抽出結果を示す。

表 2 専門用語・名詞数の条件を変えたときの抽出結果 (n=1)					
専門用語に関する制限	名詞数に関する制限	抽出数	正解数	精度	再現率
前方の名詞が専門用語	名詞数の制限なし	96	38	39.6%	82.6%
	ともに複数単語	28	14	50.0%	30.4%
	どちらかが複数単語	79	36	45.6%	78.3%
後方の名詞が専門用語	名詞数の制限なし	63	16	25.4%	34.8%
	ともに複数単語	17	8	47.1%	17.4%
	どちらかが複数単語	44	14	31.8%	30.4%
ともに専門用語	名詞数の制限なし	26	12	46.2%	26.1%
	ともに複数単語	11	6	54.5%	13.0%
	どちらかが複数単語	21	11	52.4%	23.9%
どちらかが専門用語	名詞数の制限なし	133	42	31.6%	91.3%
	ともに複数単語	34	16	47.1%	34.8%
	どちらかが複数単語	102	39	38.2%	84.8%

表 3 専門用語・名詞数の条件を変えたときの抽出結果 (n=1.2)					
専門用語の制限	名詞数の制限	抽出数	正解数	精度	再現率
前方の名詞が専門用語	名詞数の制限なし	90	38	42.2%	82.6%
	ともに複数単語	26	14	53.8%	30.4%
	どちらかが複数単語	74	36	48.6%	78.3%
後方の名詞が専門用語	名詞数の制限なし	60	15	25.0%	32.6%
	ともに複数単語	16	7	43.8%	15.2%
	どちらかが複数単語	42	13	31.0%	28.3%
ともに専門用語	名詞数の制限なし	24	11	45.8%	23.9%
	ともに複数単語	10	5	50.0%	10.9%
	どちらかが複数単語	19	10	52.6%	21.7%
どちらかが専門用語	名詞数の制限なし	126	42	33.3%	91.3%
	ともに複数単語	32	16	50.0%	34.8%
	どちらかが複数単語	97	39	40.2%	84.8%

上記の結果を踏まえ、「前方の名詞が専門用語」「前方・後方どちらかの名詞が複合名詞」の条件において (1) 式における n の値を変え、それぞれについて精度・再現率を調査した。表 4 に結果を示す。結果より、精度・再現率が総合的に高くなったのは $n = 1.2 \sim 1.5$ であった。評価実験では $n = 1.2$ を用いる。

表 4 n の値を変化させたとき精度及び再現率

n	精度 [%]	再現率 [%]
0.8	44.4	78.3
1.0	45.6	78.3
1.1	48.0	78.3
1.2	48.6	78.3
1.3	48.6	78.3
1.4	48.6	78.3
1.5	48.6	78.3
2.0	49.3	76.1
3.0	50.0	71.7
5.0	49.1	58.7

5 評価実験

5.1 概要

本手法を用いて解析を行い、有用性の評価を行った。絞り込みの条件やパラメータの値は 4 章の結果を用いて決定した。評価対象には、4 章で用いたものと同様の長野県議会会議録「令和元年 6 月定例会本会議 02 号」1 件を用いた。文字数は 72,453 文字であった。また、この中で抽出したい社会課題を人手で選定したところ 207 件が得られた。これを評価実験における正解データとして扱う。

5.2 専門用語辞書の作成

表 5 にカテゴリごとの登録語数を、表 6 に専門用語として辞書に登録されたカテゴリごとの名詞の例を示す。

表 5 作成した辞書におけるカテゴリごとの登録語数

カテゴリ	登録語数
教育	18757
健康・福祉	16445
農業	8788
林業	6905
環境	3931
観光	5402
インフラ	11422
警察	5664
広報	5986
文化	3180
災害	3633
企業・労働	6416
行政	1545

表 6 より、登録された名詞はそのカテゴリの専門用語であると考えられるものが多い。しかし、辞書の中には一部人名や特定の数値が含まれるものなどが含まれており、改善の余地がある。

5.3 社会課題の抽出

(a) 候補の抽出

会議録中の回答者の発言全てについてパターンに基づいた抽

表 6 辞書に登録された名詞の例

教育	健康福祉	農業	農業
学校	健康福祉部長	農地	自然エネルギー
教育長	医師	農業	廃棄物
生徒	医療機関	農業者	太陽光発電
教員	病院	農家	環境省
児童生徒	医療	担い手	最終処分場
教育	衛生部長	農産物	放射性物質
県教育委員会	県立病院	生産	小水力発電
子供たち	サービス	米	メガソーラー
高校	看護師	消費者	合併処理浄化槽
保護者	患者	生産者	地下水
教育委員会	養成	遊休農地	浄化槽
学び	介護	新規就農者	温室効果ガス
指導	がん	農政部	水質
市町村教育委員会	事業所	リンゴ	モニタリング
学習	治療	J A	温暖化対策
授業	障害者	中山間地域	埋め立て処分

出を行い、すべての候補を取得した。結果、859 件を取得した。この中に正解データ 207 件は全て含まれていた。

(b) 候補の絞り込み

絞り込みを行った結果を表 7 に示す。

表 7 絞り込みの結果

絞り込み前 (候補) の数	859
絞り込み後の数	289
正解データと一致した数	128
精度	44.3%
再現率	61.8%

精度が低くなった原因として、社会課題の定義を満たしていない抽出が多かった点が挙げられる。対象、あるいは目的が含まれていなかったり、曖昧であったりして 2.1 節 (1) の条件を満たさない抽出があり、これらが精度の低下につながったと考えられる。定義を満たさない、誤った抽出の例に「教員志願者確保のため」という例があった。「教員志願者確保」であれば社会課題と判断して問題ないため、抽出パターンの拡張・検討が必要である。また、他の社会課題と合わせて見ることで単体では社会課題として認められない抽出が社会課題と判断できるケースもある。具体例として、「教育」カテゴリの、

- 特別選考枠の教員採用
- 志願者確保の取り組み
- 教員採用選考への志願者数

という例では、「特別選考枠の教員採用」「教員採用選考への志願者数」は社会課題として認められるが、「志願者確保の取り組み」という抽出は単体では対象を正確に理解できず、社会課題として認められない。しかし、これら 3 つの抽出を同時に見ることで対象が「教員」であるとわかる。もともと関連のない抽出結果同士を誤って関連付けてしまう場合も想定されるが、いずれにしても可視化をしたうえでの評価方法をさらに検討する必要がある。

5.4 カテゴリの付与

抽出・絞り込みを行った社会課題 289 件に対してカテゴリの付与を行った。289 件中 288 件に正しく付与を行えた。そのうち所属からカテゴリを判断できるケース 273 件にはすべて正しく付与できた。表 8 に抽出した社会課題と付与されたカテゴリの例を示す。表中の「農作業の体験」という社会課題に対して付与された「教育」のカテゴリのように、社会課題からのみでは趣旨を詳しく理解できない場合でも、カテゴリを付与することで補足できることがわかる。

表 8 社会課題に対して付与されたカテゴリの例

抽出された社会課題	付与されたカテゴリ
農作業の体験	教育
災害拠点病院への財政的支援	健康福祉
農業の魅力発信	農業
外国人特定技能制度の活用	農業
農業労働力の確保	農業
多文化共生推進指針の改定	文化
アカマツの植栽	林業
マツタケ山の再生	林業
保安林の維持造成	林業
土砂崩落防止のため施設整備	林業

また、抽出された 289 件の中に回答者の所属からではカテゴリを特定できないケースは 16 件あった。結果を表 9 に示す。16 件中 15 件に正しく付与を行えた。正しく付与を行えなかったのは「米軍機の低空飛行」という抽出であり、「災害」カテゴリに分類された。この例の場合、どのカテゴリでも明確に適切であるとは言い切れない。よって、カテゴリのリストについても拡張・検討が必要である。

表 9 所属からカテゴリを特定できないケースでの結果

社会課題	付与されたカテゴリ
中部横断自動車道の整備促進	インフラ
都市計画の手続	インフラ
八千穂間の早期	インフラ
米軍機の低空飛行	災害
学習活動の実情	教育
不登校児童生徒の状況	教育
子供たちの希望	教育
教育委員会の方針	教育
高齢ドライバーの事故	警察
高齢ドライバーの方	警察
高齢運転者の交通安全緊急対策	警察
安全運転サポート車の普及推進	警察
交通安全協会の皆様方	警察
約 250 の関係部局	健康・福祉
社会との役割分担	教育
教育委員会の所管	教育

6 おわりに

本研究では可視化を目的とした会議録からの社会課題の抽出、及びカテゴリの付与を行った。限定されたパターンではあるが、社会課題を抽出し、カテゴリ付与を行うまでの手法を提案した。特に、回答者情報をもとにしたカテゴリ付与や専門用語辞書の作成等、一定の有用性を得られた。しかし、抽出した社会課題にはノイズが多く、実用には精度の向上が望まれる。そのために、今後の課題には前述の抽出パターンの拡張、抽出後の評価方法の見直し、カテゴリ付与の精度向上、カテゴリリストの拡充が挙げられる。

さらに、今回は長野県議会会議録のみを元データとして実験を行った。他の自治体でも同様の手法が適用できるのか、より多くの種類の会議録を用いて実験を行い、調査していく必要がある。

参 考 文 献

- [1] 木村泰知, 関根聡, 乾健太郎. 地方議会会議録の要約に向けて. 言語処理学会第 24 回年次大会発表論文集, pp. 596–599, 2018.
- [2] 乙武北斗, 高丸圭一, 内田ゆず, 木村泰知. 述語項構造に基づく地方議会会議録からの意見抽出手法の検討. 日本知能情報ファジィ学会ファジィシステムシンポジウム講演論文集, Vol. 34, pp. 222–225, 2018.
- [3] 高丸圭一, 内田ゆず, 木村泰知. 地方政治コーパスにおける都道府県議会会議録パネルデータの基礎分析. 宇都宮共和大学 シティライフ学論叢, Vol. 18, pp. 136–155, 2017.
- [4] 木村泰知, 関根聡. 主辞に基づく政治問題抽出手法. 人工知能学会論文誌, Vol. 28, No. 4, pp. 370–378, 2013.
- [5] 長野県議会 会議録検索システム. <https://nagano.gijiroku.com/voices/index.asp>. (2020 年 2 月 4 日参照).