

地方議会会議録を利用した自治体間引用関係抽出における被引用数・引用数の分析

葦原 史敏^{†,a} 木村 泰知^{‡,b} 荒木 健治^{†,c}

[†] 北海道大学大学院情報科学研究科 [‡] 小樽商科大学 商学部 社会情報学科

a) ashihara@media.eng.hokudai.ac.jp b) kimura@res.otaru-uc.ac.jp c) araki@ist.hokudai.ac.jp

概要 本研究では、他の自治体と比較して重要な自治体を発見するために、政策課題ごとに全国の地方議会でも多く引用される自治体を抽出する。また、抽出されたデータから、当該自治体が他の自治体から参考にされている傾向を示す「被引用数」と、当該自治体が他の自治体を参考にしている傾向を示す「引用数」を抽出する。そして、他の自治体から参考とされる自治体が他の自治体を参考にしている傾向を示すか明らかにするために、これらの二つの値の相関係数について分析する。

キーワード 地方議会会議録, 引用関係, 引用数, 被引用数

1 はじめに

近年、全国の約7割の自治体が、住民への情報公開に向けて、ウェブを利用して地方議会会議録（以下、会議録）を公開しており、住民は自治体の意思決定に関する議員らの議論を閲覧することができるようになっている[1]。会議録は全国の自治体で蓄積されており、利用状況に差があるブログやマイクロブログ (twitter) などの情報源と比較すると、全ての自治体が情報を公開している媒体である。そのため、特定の自治体に限らず、情報を閲覧することが可能である。一方で、自身の自治体以外の会議録を検索することは、異なる自治体を参考にし、比較対象とする上で重要であるが、同じ政治課題を抱えている自治体や、注目される取組がある自治体を調べること自体が容易ではないため、困難が伴う。

この問題に対し、本研究では自身の自治体に関連する、異なる自治体の会議録をシステムによって推薦する手法を実現することを最終目標とする。図1に概要を示す。

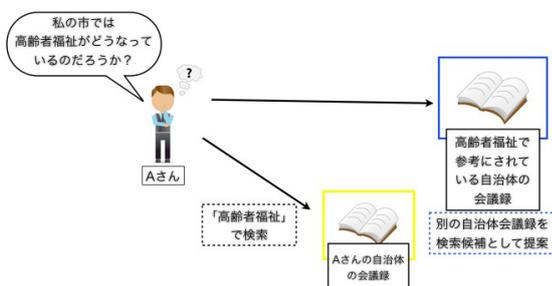


図1 システムの概要

一方で、注目される取組がある自治体を判別することは、異なる自治体から参考にされる取組を行っていることを示す発言を会議録から収集することで実現できると考えられる。異なる自治体の取組について言及する発言の例を図2に示す。

次なる森林の取組について注目すべきは企業の森、これは、和歌山県がやはり平成15年に全国に先駆けてこれを実施いたしました。

図2 異なる自治体の取組について言及する発言の例

図2は論文の引用と類似した機能を持つ発言であるため、他の自治体から引用される回数が高いほどその自治体の重要度が高いと考えられる。本研究ではこのような発言を「引用発言」と呼ぶ。本稿では、引用発言の特徴を明らかにするために、特定の政治課題に対してどのような自治体が全国の会議録でよく言及されているか調査する。調査は当該自治体から参考とされている傾向を示す「被引用数」と、当該自治体から参考にする傾向を示す「引用数」を抽出して行う。また、他の自治体から参考とされる自治体から参考にする傾向を示すかどうかについても明らかにするために、これらの値の相関係数についても分析する。

2 引用発言抽出

本章では他の自治体を引用する発言の抽出方法について述べる。

対象データは2010年の全国の会議録から[2]に示す方法で収集された434の自治体の会議録とする。2010年を対象としたのは最もデータ数が多いため、レコード数は11,681,563 (1.69GB)である。これらの会議録から自身の自治体名を除いて全国の自治体名を含む発言を取得する¹。

しかし、自治体名を含む発言を取得するだけでは、当該自治体を参考にしている意図が無い発言も含まれてしまう。例えば、札幌市をキーワードとすると、札幌市長を

Copyright is held by the author(s).

The article has been published without reviewing.

¹対象自治体名のリストは総務省の平成25年のデータを使用した。
<http://www.soumu.go.jp/denshijiti/code.html>

含む発言も抽出され、発言の焦点が自治体ではなくなることや、行政上関係のある近隣自治体の名前に言及する発言が含まれるなどの問題がある。そのため、以下の2点について対応した。

- 「～市長」など、自治体名ではなく人物を指す固有名詞などが抽出されるため、照合した自治体名の直後に「長」などの語が含まれる場合、抽出対象から除外する。
- 同一都道府県内の自治体への引用を対象外とする。

なお、前者については、形態素解析や固有表現抽出によって自治体名とマッチすることで解決できると考えられるが、対象データのファイル量が膨大であり、形態素解析等を行うと著しく処理時間が増加するため、これらの解析器を使わない方法で対応した。

また、上記の同じ自治体名を持つ都道府県の曖昧性を解消できないという問題が存在する。例えば、伊達市は北海道と福島県に存在し、当該自治体名を含む発言はどの都道府県に所属する自治体を引用しているか判別できない。この問題の解消は今後の課題である。

以上の抽出方法により、80,642 発言が抽出される。

3 調査

抽出された 80,642 発言から、特定の政治課題に対しどのような自治体が言及されているか調査する²。対象とする政治課題は、全国の自治体で政治課題となっている「高齢者福祉」、「児童福祉」、「地域活性化」の3つを対象とし、政治課題の特定は表1に示す政治課題ごとのキーワードを暫定的に設定し、キーワードのいずれかが含まれるかによって判断する。

表1 政治課題ごとのキーワード

政治課題	キーワード
高齢者福祉	高齢者, 介護, 少子化, 高齢化
児童福祉	児童, 保育
地域活性化	地域活性化, まちづくり, 住民参加

また、各政治課題ごとに当該自治体が他の自治体から参考にされる傾向を示す「被引用数」と当該自治体が他の自治体を参考にする傾向を示す「引用数」を抽出する。

表2に政治課題ごとの「被引用数」と「引用数」の上位5自治体と、二値の相関係数を示す。なお、一度の発言で複数の自治体を引用する場合は引用数を1とし、相関係数を計算する際は会議録のデータが存在しない自治体への引用発言(1,082自治体)は除外した。

4 考察

被引用自治体の上位はいずれの政治課題においても東京都とその自治体が上位となっており、政策のモデル

²ただし、引用自治体で湖南市のデータが江南市のデータとして誤って登録されていたため、データから除外した。

表2 高齢者福祉における被引用数の高い上位5自治体

自治体名	被引用数	自治体名	引用数
高齢者福祉 (相関係数: .101)			
稲城市	162(7.04)	福島市	36(1.94)
東京都	121(5.26)	目黒区	32(1.72)
足立区	97(4.22)	坂戸市	26(1.40)
札幌市	65(2.83)	石巻市	25(1.35)
群馬県	46(2.00)	川崎市	25(1.35)
児童福祉 (相関係数: .277)			
東京都	123(6.89)	西宮市	36(2.41)
大阪府	80 (4.48)	新宿区	34(2.28)
横浜市	60 (3.36)	川崎市	29(1.94)
北区	44 (2.47)	横浜市	26(1.74)
江戸川区	37(2.07)	堺市	22(1.47)
地域活性化 (相関係数: .100)			
東京都	30(2.52)	印西市	16(1.80)
二セコ町	19(1.60)	川崎市	15(1.69)
富山市	18(1.51)	盛岡市	13(1.46)
金沢市	15(1.26)	甲賀市	13(1.46)
横浜市	15(1.26)	福島市	12(1.35)

注) () 内は%を表す。

ケースとして東京都の自治体が参考にされる傾向が理解できる。引用数には共通の特徴は見られないが、被引用数と比較して特定の自治体に偏る傾向は見られない。また、被引用数と引用数の間に相関は見られなかった。したがって、他の自治体から参考にされる自治体は必ずしも他の自治体を参考にする傾向にないことが理解できる。

5 おわりに

本研究では、自身の自治体に関連する異なる自治体の会議録を推薦する手法を実現するために、引用発言の抽出とその特徴の調査について検討した。特に、特定の政治課題について「被引用数」、「引用数」、その二つの相関係数を調査した。その結果、被引用数の上位に特徴が確認され、また被引用数と引用数の間の相関係数から、他の自治体から参考にされる自治体は必ずしも他の自治体を参考にする傾向にないことが確認された。

今後の課題として、発言の政治課題の特定をキーワードによらない適切な分類方法について検討する必要がある。また、時系列による変化など、他の分析軸についても検討する予定である。

参考文献

- [1] 高丸圭一, 木村泰知, 渋木英潔: 全国の市町村議会 会議録のウェブ公開とデータ提供の状況, 宇都宮共和大学都市経済研究年報, 第11号, pp.47-72, 2012.
- [2] 菅原晃平, 大城卓, 斎藤誠, 永井隆広, 渋木英潔, 木村泰知, 森辰則, 地方議会会議録コーパスの拡充における問題点の分析と対処, 言語処理学会第18回年次大会発表論文集, pp.1-15, 2012.