HOKKAIDO UNIVERSITY

Title	地方議会会議録を対象とした意見抽出に関する研究［論文内容及び審査の要旨］
Author(s)	葦原, 史敏
Degree Grantor	北海道大学
Degree Name	博士(情報科学)
Dissertation Number	甲第12637号
Issue Date	2017-03-23
Doc URL	https://hdl.handle.net/2115/65717
Rights(URL)	https://creativecommons.org/licenses/by-nc-sa/2.1/jp/
Type	doctoral thesis
File Information	Fumitoshi_Ashihara_abstract.pdf, 論文内容の要旨

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

学　位　論　文　内　容　の　要　旨

博士の専攻分野の名称　　博士（情報科学）　　氏名　葦原 史敏

学　位　論　文　題　名

地方議会会議録を対象とした意見抽出に関する研究
（A Study on Opinion Retrieval from Regional Assembly Minutes）

TV や新聞のように時間や紙面に限りがあるメディアで取り上げられる政治情報は, 国政に関する内容が中心的であり, これに比べて地方政治に関する情報は少ない. 議員活動についても同様で, 地方議会議員は国会議員と同様に住民による選挙によって選ばれ, かつ, 国政よりも身近な存在であるべきであるにもかかわらず, その活動に関する認知度は国会議員よりも低い. また, 平成の大合併により, 各自治体の広域化が進み, 各議員の活動を把握することが困難となっている.

そのような状況において, 近年全国の約 7 割の自治体では, 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.

しかしながら, 会議録は議員の発言を書き起こした話し言葉であり, 要約されていない読みづらい膨大なテキストデータであるため, ウェブ上に公開しただけでは, 読もうと考える住民がほとんどいない. このため, 会議録は公開されているにも関わらず住民によって閲覧される機会が少なく, 情報提供形態を工夫し, 会議録の情報を提供することが望まれている. そこで本研究では, 住民や行政関係者が全国の議会で議論される政治課題や政治課題を解決する方法の把握を容易にするために, 自治体に対する議員の要望意見, 他自治体の政策を引用する議員の意見の抽出を目指す.

本学位論文では, まず議員の要望意見を抽出する手法を述べる. 地方議会会議録の特徴としては, 首長や議員の発言を書き起こした話し言葉であり, 一文が長くなる点が挙げられる. そのため, 要望を抽出する単位は, 地方議会会議録の特徴を踏まえ, 文内の要望を含む箇所のみを抽出するために「節単位」としている. 本手法は, 機械学習を用いて節単位で要望の有無を判定しており, 機械学習の素性として, 後続の節のモダリティ表現の出現を捉えるパターンや, 節境界ラベルなどの構文的情報を利用している. 評価実験では適合率, 再現率,F 値を用いた評価を行い, 節内の bag-of-words のみを用いた場合と, 全ての素性を用いた場合を比較し, クローズドデータの F 値で 33.6 ポイント, オープンデータの F 値で 24.3 ポイントの向上が確認された. 特に, 他の素性に節境界ラベルを追加した際に, 適合率を低下させることなく再現率を向上することが確認された.

次に, 他自治体の政策を引用する議員の意見の抽出手法について述べる. 他自治体の政策を引用する議員の意見を抽出するためには, 発言中に自治体名が存在し, かつ当該自治体の政治的取組を含むものを抽出する必要がある. しかしながら, 発言中の自治体名には曖昧性のある名称が存在する問題がある. 地名の曖昧性解消は, 住所データから曖昧性が存在する地名を収集し, 発言中の地名それぞれについて正しい地名候補を決定することで行う. 正しい地名候補の決定は, 上位地名階層を見つけるパターンマッチ, 発言が行われた会議録の地理情報, 発言から自動抽出した地名の地理情報に基づいて行った. 実験では,3 つの手順を組み合わせることで 0.895 の高い適合率で曖昧性を解消できることがわかった. さらに, 地名の曖昧性解消手法を組み合わせ, 他の自治体の政策や取組を引用し, 参考にする意図を持つ発言の抽出を試みた. まず正解データ作成が可能か検証するために, 出典

議会と異なる自治体名を含む文を抽出し, これらの文に対して 3 名の評価者で対象の文と前後の文から, 政策や取組を引用し, 政策を参考にする意図を持つ発言かを評定し一致率を評価した. その結果,Cohen の kappa 係数として 0.44〜0.58 という中程度の一致が得られた. また, 得られた正解データを用いて機械学習によって政策や取組を引用するものか判別する分類器も構築し, 分類精度について検証した. 提案手法では bag-of-words, ルールベース, 地理上の距離,85 歳以上の人口類似度を素性として用いた. 実験では,bag-of-words のみの素性を用いた場合と比較して,F 値が 0.016 ポイント向上することが確認された.

今後の研究課題としては, 自治体に対する議員の要望意見の抽出, 他自治体の政策を引用する議員の意見の抽出, またそれに用いる自治体名の曖昧性解消の精度を改善することがあげられる. また, 各自治体の議会における要望意見の抽出による会議録の読解支援システムの構築や, 高頻度で引用される自治体の可視化システムを構築することも今後の課題である.