

テキストマイニングを用いた議事録の 可視化手法に関する検討 —廃棄物分野を事例として—

小山 文敬¹・藤山 淳史²・石井 一英³・佐藤 昌宏⁴

¹非会員 北海道大学修士課程 大学院工学院 (〒060-8628 札幌市北区北13条西8丁目)

E-mail:fumi0403@eis.hokudai.ac.jp

²正会員 北九州市立大学講師 環境技術研究所 (〒808-0135 北九州市若松区ひびきの1-1)

E-mail:fujiyama@kitakyu-u.ac.jp

³正会員 北海道大学教授 大学院工学院 (〒060-8628 札幌市北区北13条西8丁目)

E-mail:k-ishii@eng.hokudai.ac.jp

⁴正会員 北海道大学助教 大学院工学研究院 (〒060-8628 札幌市北区北13条西8丁目)

E-mail:satomasahiro@eng.hokudai.ac.jp

行政主体の廃棄物分野の委員会では合意形成が重要であり、多数の利害関係者存在下での意思決定プロセスの透明化と情報公開が欠かせない。そこで、廃棄物分野の全文公開された行政主体委員会の議事録を対象として、テキストマイニングによる議事録の客観的な可視化を試みた。とくに本研究では、青森・岩手県境不法投棄事案を事例としてとりあげ、議事録のテキストマイニング手法を提案し、適用した。具体的には、不要語の選定、重要語の抽出を行い、さらに各委員が主張・同調表現を使用した回数から議論内容に対して批判的な意見を持つかを分析した。これらの分析により、委員の議論内容への態度という観点から、議事録を客観的に可視化できた。

Key Words: waste management, minutes, text mining, visualization of committee's attitude

1. はじめに

議論の客観的な分析手法として、議事録のテキストマイニングという手法がある。テキストマイニングとは一般に、テキストデータを数値データに処理した後、統計的手法を適用するデータマイニングの一種である。議事録のテキストマイニングを行った既往研究としては、岩見らによる「公共事業計画策定過程の議事録分析による意見協調・対立関係把握のための分析手法の開発」¹⁾がある。この研究では淀川流域の河川事業をテーマに、委員会で議論されたテーマの把握、委員間の関係の可視化が行われている。また、増田による「地方議会の議事録に関するテキストマイニング分析—高崎市議会を事例として—」²⁾では、地方行政を対象として課題の把握、議会の活動調査が行われている。

その一方で廃棄物分野において議事録のテキストマイ

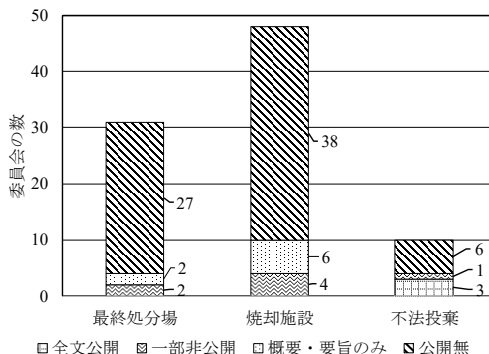


図-1 廃棄物分野の委員会等の公開状況

ニングが行われた研究例は少ない。CiNiiの論文検索 (<https://ci.nii.ac.jp/>, 2018年1月31日時点)³⁾を行ったところ、「廃棄物+テキストマイニング」では6件が検索されたが、いずれも廃棄物分野を題材として、アンケートの自由記述文を解析した研究であり、議事録を対象とした研究はなかった。廃棄物分野の委員会の公開状況を図-1に示す。なお、最終処分場と焼却施設は環境省の平成27年度の廃棄物処理実態調査⁴⁾の中から供用開始が5年以内の施設(全国)を抽出し、各施設に関連する委員会(委員会の内容は長期包括や処理方式、候補地選定など様々である)の公開状況を各HPで調べ、整理した内容である。不法投棄については「不法投棄のない循環型社会づくりー不法投棄対策のアーカイブ化ー」⁵⁾より10事案(香川県豊島事案, 青森・岩手県境事案, 秋田県能代市事案, 三重県桑名市事案, 福井県敦賀市事案, 宮城県村田町事案, 横浜市戸塚事案, 岐阜市北部地区事案, 福島県いわき市事案, 三重県四日市事案)を抽出し、各HPで調べ、整理した内容である。

その中でも今回は全文公開されている委員会の議事録を対象に分析を行った。これは、廃棄物分野の委員会議事録の公開が少ない現状において、他分野での既往研究と同様に議事録を客観的に可視化できる方法を提案することができれば、今後の廃棄物分野の委員会議事録公開に貢献できると考えたためである。また、廃棄物分野の委員会には最終処分場や焼却施設の候補地選定・整備、不法投棄現場の修復など様々なものがあり、これらの委員会には行政職員や専門家だけでなく、住民が参加することもあり多様な価値観や背景を持った利害関係者が混在することになる。このような委員会において、全関係者の合意を得た内容での意思決定を行うためには、少数意見等をも組み込んだ内容とすることは勿論、意思決定を行う際のプロセスの透明化とその情報公開を行うことは重要なポイントとなる。本研究で提案する委員会議事録のテキストマイニング解析の結果と議事録本文を併用して読むことで、関係者の意思決定に必要な情報の理解を促進することに貢献できると考えたためである。

そこで本研究では廃棄物分野の委員会議事録を対象として、テキストマイニングを適用することで、議事録を客観的に可視化する手法の開発を試みることを目的とする。

2. 分析手法の提案

(1) 議事録の分解

本研究ではテキストマイニング分析を行うためのソフ

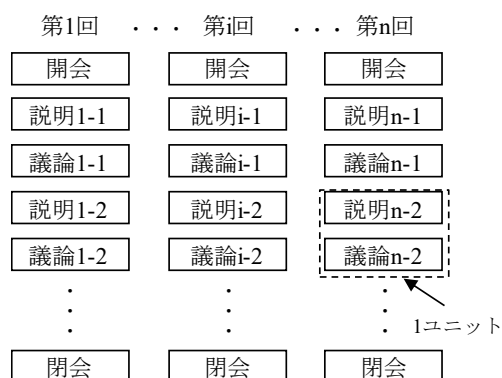


図-2 委員会を構成するユニットの構造

トウェアはTinyTextMiner (以下, ttm)⁶⁾を採用し、形態素解析を行った。なお、解析を行う際は漢字・カタカナから成る2字以上を対象とし、人名は除外した。

図-2のように、複数回の委員会から構成されている場合の議事録は、各回でまとめられていることが多い。また、議事録は開会挨拶等に始まり、事務局の資料説明部分と委員間の議論部分が議事毎に繰り返され、最後に閉会挨拶等で終わるように構成されていることが通常である。このような議事録にテキストマイニングを適用する際には、解析を行いたい内容が議論されている箇所に適用範囲を絞ることが重要である。つまり本研究のように委員の議事に対する発言に重きを置いて分析を行う場合には、開会・閉会挨拶や資料説明部分には委員間のやり取りは無く、これらは不要な部分となる。本研究では、議事録を「開会」と「閉会」、「事務局の資料説明部分」、「委員間で行われる議論部分」に分け、資料説明部分と議論部分を議事毎にユニットとしてまとめた。そのうえで、第1回委員会の1番目のユニットに対して事務局の資料説明部分には「説明1-1」、委員間の議論部分には「議論1-1」とナンバリングを行った。

(2) 不要語の選定

テキストマイニングにおいて解析に不要な単語(以下、不要語)を除くことはその後の分析結果に大きく影響を与える。本研究において不要語とは、議事録における議事の進行や質疑応答等に用いられる議事内容とは直接的に関係がないと考えられる語とした。もし事前に不要語を指定していなければ、形態素解析の抽出結果に不要語に該当する語が多く抽出されると議事内容にとって重要な語が選べれないといった問題や、抽出語を用いた統計的な分析において議事内容を捉えた分析結果が得られないといった問題が生じることになる。こういった問題が起こることを防ぐため、議事内容の分析に移る前に不要語を選定し、あらかじめ取り除く必要がある。

今回、不要語の選定に当たっては、式(1)に示す変動係数法を用いた⁷⁾。

$$CV = \frac{\sigma_{DF}}{\overline{DF}} \quad (1)$$

CV: 変動係数

DF: 第*i*回委員会を対象語が出現する文章の数

$\overline{DF}$: 第*i*回委員会での DF の平均値

σ_{DF} : 第*i*回委員会での DF の標準偏差

各語のCV値を計算し、CV値が委員会の各回で下位10%に含まれる語を不要語と設定した。つまり、委員会の各回の議事録の中でCV値のばらつきの小さい語(議事の進行や質疑応答に用いられるようななどのユニットにも出現する語)が不要語として選定されることになる。一方、専門用語のように限られたユニットにしか登場しない語はCV値が大きくなる。本来、この変動係数法は議論中に用いられた専門用語などCV値の大きな語を抽出する手法であるが、本研究では変動係数の小さい語を直接議論の内容に関係のない語とみなすことによって不要語選定の指標として用いた。

(3) 重要語の抽出

重要語とはその議論部分の特徴を捉えた語であり、後に行う分析(重要語の出現頻度などを数値データとして扱う)の最小単位とする。文章から得られる全ての語をデータとして用いることも可能であるが、そこから有用な情報を得ることは一般的に難しい。そのため、直接内容とは関係のない語を不要語として省いたり、議論の特徴を捉えていると考えられる重要語に重み付けを行ったりしない限りは、その文章を読む以上の情報を手に入れることは難しい。

そこで本研究では分析前の処理として、岩見らが提案しているDFIMF⁹⁾という重み付けの指標を改良して、新たに重要語抽出の指標DFIUFを提案する。なお、DFIMFのDFとは語が出現した段落数(Document Frequency)を、IMFとは語が出現した委員会の偏り(IMF:Inverse Minutes Frequency)を表しており、DFIUFのUは前述のユニットを表している。つまりDFIUFでは、対象語が出現する委員会の数ではなく、対象語が出現するユニットの数を扱うことにより、第*i*回委員会のなかで一部に集中して出現する語を評価することを可能とした。DFIUFの算出式を式(2)に示す。

$$DFIUF = DF \times \log \frac{U}{UF} \quad (2)$$

UF: 第*i*回委員会を対象語の出現するユニット数

U: 第*i*回委員会の全ユニット数

今回、提案する重要語抽出の手順を次に示す。委員会の各回でユニットのみを統合したデータから出現文章数の上位1000語を抽出し、抽出された語に式(2)を適用した。適用した結果をもとに、DFIUF値がそのユニットで最低値以外のものを重要語とし、DFIUF値が0のものは重要語には含めないこととした。なおこの手法は、開催回数が少ない委員会であっても、議事録の議論部分をユニットで細かく区切ることができる場合には適用可能であると考えられる。

(4) 委員の態度の分析

各委員の各ユニットでの「発言回数」と、これまでの分析によって得られた各委員が言及した「重要語数」、後述する各委員が用いた「主張・同調表現回数」の3つの指標を用いて、委員の議論への態度を分析し、委員の態度という観点から議事録の客観的な可視化を試みた。

a) 委員の発言回数による分析

委員の発言回数とは、1つのユニットにおいて各委員が発言した回数のことである。具体的にはユニットごとに各委員の発言回数を数え、比較を行う。この発言回数は委員の議論への積極性や参加度合を表していると考えられる。例えば、1人の委員があるユニットにおいて発言回数は多いが、別のユニットにおいては発言回数が少ないとき、その委員があるユニットにおいては別のユニットよりも参加度合いが高いことを意味している。

b) 委員の言及した重要語数による分析

委員の言及した重要語数とは1つのユニットにおいて各委員が言及した重要語の数(重要語が言及された回数ではなく、重要語の単語数)のことである。例えば、「シミュレーション」という重要語があった場合、それを2回発言している委員がいたとしても、言及した重要語数は1単語とカウントされる。重要語として指定された単語は委員の専門を表現するような技術的な単語が多く指定されており、この言及した重要語数は委員の技術的関心を表していると考えられる。より多くの種類の重要語に言及している委員はそのユニットで自らの専門分野に基づいた発言をしていると考えられるからである。発言回数による分析と同様に、ユニットごとで比較を行う。

c) 委員の主張・同調回数による分析

本研究において主張・同調表現とは、「私は」、「わたしは」、「私自身は」を主張表現とし、「私も」、「わたしも」、「私自身も」を同調表現とした。委員会のような議論の場において批判的な意見を発言するときは否定的な表現ではなく、このように主語が「私」など委員自身を指すものが置かれるときに主張・同調表現がなさ

れていることが多い傾向がみられたため、これらと比較することにより議論内容に対する批判的な姿勢の有無等を把握できると考えた。具体的には、1つのユニットのなかで同調表現の使用回数が0回で、かつ、主張表現の使用回数が1回以上の場合に、批判的な意見を持つ委員として指定した。

d) 委員の方針への態度

これまで分析した各ユニットにおける各委員の発言回数、言及した重要語数、主張・同調回数の結果をもとに委員の態度を参加度合い、技術的関心、批判的意見の3つの観点から可視化を行う。

3. 分析手法の適用

(1) 対象とした委員会

本研究では第1章で示した廃棄物分野の委員会のうち、全文公開されているものの中から、「青森・岩手県境不法投棄事案に係る技術部会」の議事録（全5回分）を対象に分析を行った。同技術部会とは、青森県と岩手県の県境で発覚した青森・岩手県境不法投棄事案において、両県の合同検討委員会の下部組織として修復対策について議論するために設置された委員会（全5回）である。

(2) 議事録の分解

対象とした議事録は委員会の各回とも冒頭に開会部分があった後、各議事に対して事務局が資料を説明している部分と、その内容に対して委員間で議論が行われている部分が繰り返され（交互とは限らない）ており、最後に閉会部分という内容で構成されている。第2章で述べた通り、まず初めに対象とした議事録の分解を行った。なお、今回対象とした議事録は、事務局の資料説明部分や委員間の議論部分に明確な区切りはなく、連続した文章になっていたため、各部分の区切りの基準を次に述べる通り設定することで議事録の前処理を行った。まず開会部分については、開会挨拶から委員紹介などを経て部会長が進行を始める（話し始める）までとした。次に、事務局の資料説明部分については、事務局が資料の説明を始めてから「以上」と最初に述べたところまでとした。委員間の議論部分については、部会長が開会直後に進行を始めてから事務局に資料の説明を依頼するまで、または、事務局の説明直後から部会長が次の資料説明を事務局に依頼するまでとした。その結果、議論部分が分解され、第1回は6個、第2回は5個、第3回は7個、第4回は11個、第5回は8個のユニットに分かれた。以下、前述のナンバリングに従い、第*i*回委員会の*k*個目のユ

ニットを議論 *i-k* と指定した。

ここで対象とした議事録の特徴について述べる。対象とした議事録では青森県事務局と岩手県事務局が別々に資料を作成していたことで（一部、調査結果などは途中から両県で統一して資料が作成された）、県ごとに分かれて資料が説明されている。これは、委員会での議論後の最終的な方針として青森県は全量撤去を基本に汚染拡散防止を優先する方針（以下、方針1）を採用しているのに対して、岩手県は早期全量撤去が最大の汚染拡散防止であるとする方針（以下、方針2）を採用していることに由来する。修復に向けた両県の基本的な方針が異なっていたことによって、委員間の議論部分においても各委員の意見や態度に差が表れると考えられたことから、本研究ではこの差に着目した分析を行った。

(3) 不要語選定の適用

対象とした議事録に第3章で述べた不要語の選定手法を実際に適用する。各回の議論部分のみのデータを出現する文章の数 DF が多い順に上位1000語までの名詞を *ttm* で抽出し、その中から不要語を選定した。この作業を第1回から第5回までの委員会について行った。例えば、第1回委員会を解析した際には、「部分」や「廃棄物」、「撤去」「非常」、「場合」などの語が不要語として選定された。

(4) 重要語抽出の適用

対象とした議事録に第2章で述べた重要語の抽出手法を実際に適用する。各回の議論部分のみのデータを出現する文章の数 DF が多い順に上位1000語までの名詞を *ttm* で抽出する。同時に *ttm* の機能によって選定した不要語を回ごとに取り除いた。各ユニットに出現する重要語を、その議論部分の重要語とした。これは他のユニットにおいて選ばれた語と重複する語があっても良いものとした。この作業を第1回から第5回までの委員会について行った。例えば、議論1-1を解析した際には、「遮水」や「如何」、「水処理施設」、「対策」、「緊急」などの語が重要語として選定された。

(5) 委員の態度の分析

対象とした議事録に第3章で述べた委員の態度の分析手法を実際に適用する。

a) 委員の発言回数による分析

議論部分ごとに各委員の発言回数をまとめた。議論2-5、議論3-3、議論5-8に関しては両県事務局による資料説明後のユニットとなるため、今回は分析から除外した。ユニットごとに文章の長さが異なるため、それに伴い委員

の発言回数も変動する。そのため、本解析では各ユニットでの各委員の発言回数の平均と標準偏差を用いて標準化を行った。標準化した値が大きいものほど、比較的その委員の発言回数が多かったことを示している。

さらに青森県の HP に公開されている議事次第より「除去計画基本方針」と記されている議事と対応するユニットのみをレーダーチャートで図-3 に示した。なお、青森県の事務局説明後では議論 1-4, 議論 2-4, 議論 4-9, 議論 4-10, 議論 5-3, 議論 5-4, 議論 5-5 が相当し、岩手県の事務局説明後では議論 1-5, 議論 4-7, 議論 4-8, 議論 5-6, 議論 5-7 が相当している。委員の発言回数は議論への参加度合を表すものと考えられるため、表-1 の通り、両方針の説明後で標準化された発言回数の最大値の差が平均値以上の委員を参加度合いの高い委員と指定した。

表-1 標準化された発言回数の最大値とその差

委員名	方針1 説明後の 最大値	方針2 説明後の 最大値	差	参加度合 いの偏り
部会長 A	1.579	1.696	0.118	
委員 B	2.916	-0.470	3.387	方針1
委員 C	2.063	0.252	1.811	方針1
委員 D	2.765	0.289	2.476	方針1
委員 E	1.955	-0.178	2.132	方針1
委員 F	2.100	2.100	0.000	
委員 G	1.999	1.466	0.533	
委員 H	1.403	2.067	0.664	
委員 I	0.225	2.535	2.311	方針2
平均			1.402	

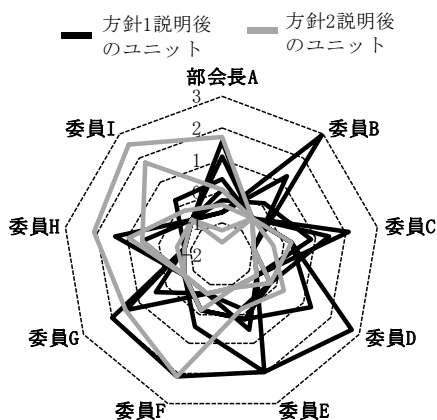


図-3 除去計画基本方針に関わる議論における委員の発言回数

したがって、方針1の説明後により参加度合いの高い委員は委員 B, C, D, E であり、方針2の説明後に参加度合いの高い委員が委員 I であることがわかった。両方針で差が見られない委員は部会長 A, 委員 F, G, H であった。

b) 委員の言及した重要語数による分析

ユニットごとに各委員の言及した重要語数をカウントする。議論 2-5, 議論 3-3, 議論 5-8 に関しては両県事務局による資料説明後のユニットとなるため、今回は分析から除外した。以下、委員会の発言回数による分析と同様に分析を行った結果を図-4 及び表-2 に示した。方針1の説明後により技術的関心の高い委員は委員 B, E, G であり、方針2の説明後に参加度合いの高い委員が委員 I であることがわかった。両方針で差が見られない委員

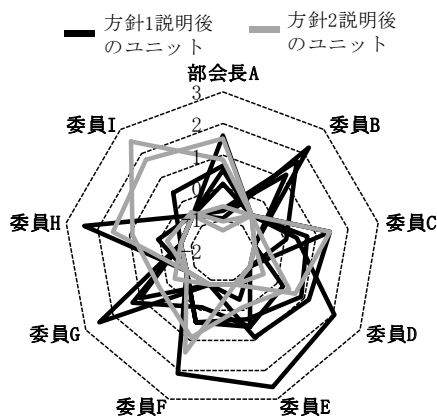


図-4 除去計画基本方針に関わる議論における委員の言及した重要語数

表-2 標準化された言及した重要語数の最大値とその差

委員名	方針1 説明後の 最大値	方針2 説明後の 最大値	差	参加度合 いの偏り
部会長 A	1.629	1.551	0.079	
委員 B	2.248	-0.562	2.810	方針1
委員 C	1.431	1.431	0.000	
委員 D	2.065	0.662	1.402	
委員 E	2.569	-0.106	2.675	方針1
委員 F	2.141	1.393	0.748	
委員 G	2.473	0.284	2.188	方針1
委員 H	2.403	1.527	0.876	
委員 I	0.420	2.512	2.092	方針2
平均			1.430	

表-3 方針1説明後の議論部分における委員の主張・同調表現回数

委員名	主張表現を用いた回数	同調表現を用いた回数
部会長 A	1	2
委員 B	4	0
委員 C	0	1
委員 D	1	1
委員 E	0	1
委員 F	6	2
委員 H	2	1
委員 I	1	0

表-4 方針2説明後の議論部分における委員の主張・同調表現回数

委員名	主張表現を用いた回数	同調表現を用いた回数
部会長 A	2	1
委員 F	2	0
委員 H	8	0
委員 I	1	0

は部会長 A、委員 C、D、F、H であった。

c) 委員の主張・同調回数による分析

主張・同調表現の使用回数の比較は批判的な意見を持つ委員を評価できると考えた。表-3 及び表-4 において、同調表現を使用せず主張表現のみを使用した委員をその方針への批判的な意見を持つ委員とした。よって、方針1 説明後に批判的な意見を残したと考えられる委員は委員 B であり、方針2 の説明後に批判的な意見を残したと考えられる委員が委員 F、H であった。両方針で差が見られない委員は部会長 A、委員 C、D、E、G、I であった。委員 I は両方針に対して同調表現を使用せず主張表現を使用しているが、回数が同じで判断できなかったため差が見られない委員とした。

d) 委員の方針への態度

これまで分析した各ユニットにおける委員の発言回数、言及した重要語数、主張・同調回数の結果を表-5 にまとめ、委員の参加度合いと技術的関心、批判的意見から総合的に判断し、委員の方針への態度を特徴づけた。これにより各委員の2つの修復方針に対する態度の観点から議事録を客観的に可視化することができた。部会長 A はいずれの方針への態度も表れなかった。これは部会長という立場上、中立的な役割を担っていたためではないかと考えられる。委員 B は最も方針1 に対して批判的であったと考えられる。委員 C、D は方針1 の議論で積極的

表-5 表委員の方針への態度による特徴づけ

委員名	参加度合い	技術的方針	批判的意見
部会長A			
委員B	方針1	方針1	方針1
委員C	方針1		
委員D	方針1		
委員E	方針1	方針1	
委員F			方針2
委員G		方針1	
委員H			方針2
委員I	方針2	方針2	

に発言したと考えられる。委員 E、I は方針1・2 に対して批判的意見は残していないが自らの専門分野に基づいて多く発言していたことがわかる。委員 F、H は、発言回数は目立たないが方針2 へ批判的意見を残したと考えられる。委員 G は方針1 に専門的な意見をしたと考えられる。

4. 結論

第2章で提案した分析手法を第3章において実際の議事録に適用した。対象とした議事録において、委員の発言回数、DFIUFによる重要語、委員の主張・同調回数について分析を行った。分析によって得られた委員の議論への参加度合い・技術的関心・批判的意見から、方針への態度を客観的に可視化した。提案した分析手法によって、議事録を読むだけでは読み取りづらい情報をテキストマイニングによって得られる可能性を示した。対象とした議事録では、両県で方針が違うこと、両県が別々に資料説明を行っていたことが特徴として挙げられ、これを利用することで分析を行うことができた。したがって、同様の形態の委員会議事録については、この手法が適用できると考えられる。

今後の課題としては、議事録の内容を直接分析する手法が、委員の主張・同調回数による分析のみに限られている点が挙げられる。また、委員の主張・同調についても、主語が「私」であるものだけでは限らず、それ以外の表現の主張・同調については今回の分析で考慮されていないため、今後更なる工夫が必要であると考えられる。再現性という点では、他の議事録についても今回提案した分析手法を適用し、より多くの議事録に対応できるよう改良していくことが必要である。一方で、廃棄物分野の現状として、委員会の議事録等が公開されている事例は少なく、議事録が全文公開されているものはさらに少

ない。そこで、今回提案したようなテキストマイニングによる議事録の有効利用法を今後発展させていくことが、議事録の全文公開のインセンティブにつながり情報公開を促進するきっかけになればと考えている。

参考文献

- 1) 岩見麻子・大野智彦・木村道徳・井手慎司:公共事業策定過程の議事録分析による意見の協調・対立関係把握のための分析手法の開発, 土木学会論文集 G (環境), Vol.68, No.6 (環境システム研究論文集 第40巻), II_411-II_418, 2012.
- 2) 増田正: 地方議会の議事録に関するテキストマイニング分析-高崎市議会を事例として-, 地域政策研究=Studies of regional policy 15(1), 17-31, 2012.
- 3) CiNii Articles, <<https://ci.nii.ac.jp/>>, 2018. 8. 11 参照.
- 4) 環境省:平成 27 年度一般廃棄物処理実態調査結果, <http://www.env.go.jp/recycle/waste_tech/ippan/h27/index.html>, 2018. 8. 14 参照.
- 5) 古市徹・西則雄:不法投棄のない循環型社会づくり—不法投棄対策のアーカイブス化—, 環境新聞社, 2009.
- 6) 松村真宏, 三浦麻:TTM: TinyTextMiner β version, <<http://mtmr.jp/ttm/>>, 2018. 8. 11 参照.
- 7) 岩見麻子・大野智彦・木村道徳・井手慎司:公共事業策定過程の議事録分析のための変動係数法を用いた対象語選定手法の開発, 環境情報科学学術研究論文集 25, pp.55-60, 2011.

(2018. 8. 24 受付)

STUDY OF VISUALIZATION METHOD OF CONFERENCE MINUTES USING TEXT MINING - A CASE STUDY OF WASTE MANAGEMENT FIELD -

Fumitaka KOYAMA, Atsushi FUJIYAMA, Kazuei ISHII and Masahiro SATO

Consensus building is important for administrative committees in waste management field. The decision-making process under presence of many stakeholders should be transparent and disclosure of all information is required. Therefore, we attempted systematic visualization of minutes of committees by text mining for the minutes of the administrative subject committee, which was published in the whole sentence in waste field. Especially, in this study, we proposed a text mining method for minutes and applied it to a case of Aomori · Iwate prefecture illegal dumping. Specifically, we selected unnecessary words, extracted important words, and further analyzed whether each member had critical opinions on the content of the issues from analyzing the number of times each member used the asserting or agreeing statement. Through these analyzes, we were able to systematically visualize the minutes from the viewpoint of the committee's attitude to the discussion.