

市議会会議録と住民ブログとのマッチングのための共通カテゴリの付与

Annotation of Common Categories For Matching
Between Minutes of Municipal Assemblies and Inhabitants Blog木村泰知^{*1}

Yasutomo KIMURA

渋谷英潔^{*2}

Hideyuki SHIBUKI

^{*1} 小樽商科大学

Otaru University of Commerce

^{*2} 横浜国立大学

Yokohama National University

Our aim is to consider the possibilities of information matching between city assemblymen's behavior and inhabitant's opinion using WWW. This paper describes the result which analyzed both of minutes of Otaru city assembly and Otaru's inhabitant blog using political categories. And we confirmed the ratio of annotated political categories.

1. はじめに

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。このような住民に提供される地方政治の情報、特に地方議会議員に関する情報量の不足を解決するための方法の一つとして、ウェブ上の情報を有効に利用することが考えられる。地方議会や議員側が発信するウェブ上の情報には、自治体が公開する会議録や、各政党や議員自らが運営するHPやブログなどが存在している。しかしながら、これらの情報は不特定多数を対象として発信されたものであるため、単純に収集して提示するだけでは個々の利用者にとって必要な情報を多く含むこととなる。それゆえ、利用者である住民の興味や関心に合った情報のみを整理して提示することが必要となる。

一方で、利用者が自分の政治的な興味や関心が何であるかを明確に認識できていない場合があり、キーワードなどによる明示が困難な場合がある。このような状況において、利用者個人の日常生活における不満や要望などから政治的関心を推測することができたならば利用者にとって必要な情報のみを提示することが可能となる。個人が発信するウェブ情報には、個人のホームページ、ブログ、SNS、掲示板、チャットなどによるものがある。この中でもブログは近年、一般市民に急速に普及した情報発信手段であり、2008年までにその開設数は1,690万件を超えているといわれている。ブログには日々の生活で感じたことをありのままに書かれていることが多く、例えば、「今日、海水浴に行ったらゴミだらけで不愉快だった。もう行きたくない。」といった記事が書かれることは十分に考えられる。我々は、そのような記述から、ゴミの収集方法や環境美化、観光客の誘致といった政治的課題を明らかに、会議録などの地方議会や議員側が発信する情報と対応付けることで、利用者である住民の興味や関心に合った情報を提示することを目的とした研究を行ってきた[1][2][3]。

しかしながら、政治用語などの硬い表現が含まれる会議録の

文章と、口語などの軟らかい表現が含まれるブログ記事とを、表層的な表現のみを手掛かりとして結び付けることは困難な課題である。そこで、文献[1]では、分類語彙表を用いて表層表現を拡張することにより、小樽市の市議会会議録とブログ記事のマッチングを試みた。また、文献[2]ではブログ記事に含まれる政治的課題を分析し、記事中のどのような表現が手掛かりとなっているかを調査した。これらの結果、手掛かりとなる表現を一般的な概念体系によって拡張するだけではマッチングに限界があることが判明した。そこで、我々は地方政治の問題に特化した概念体系を用いて会議録とブログの記述を共通の概念に抽象化することで、この問題を解決することを考え、共通概念を表すための政治的カテゴリの提案を行った[3]。本稿では、この政治的カテゴリを用いて、小樽市の市議会会議録1年分と小樽住民のブログ40サイトに対して人手による注釈付けを行い、我々が提案した政治的カテゴリが実際のウェブ文書に出現しているか、また、その分布に偏りがあるかを調査する。加えて、調査結果から両者のマッチングに向けた方法を検討する。

本稿の構成は以下の通りである。2.では、我々が提案する政治的カテゴリの作成方法および内容に関して記述する。3.と4.では、市議会会議録と住民ブログに対して、それぞれ政治的カテゴリを付与した結果をクロス分析する。5.では分析結果を基に効果的にマッチングを行う方法を検討する。6.は結論である。

2. 政治的カテゴリ

政治的カテゴリは、地方議員と住民を結びつけるためのものであり、地方議員の活動を抽象化するために利用する。既存の政治に関するカテゴリの多くは国政に関する内容が多く、地方議会の内容として相応しくないカテゴリが含まれるため、我々は小樽市、帯広市、函館市、釧路市の4市を対象とした予備調査を行った。その結果、議題を区分するために存在する委員会体系が4市に共通していることが確認された。そこで、我々は、委員会体系が地方政治における基本となる概念体系であると仮定し、最も細目化されている帯広市の市議会における常任委員会とその所管事項の名称をもとに基本となる概念体系を作成することにした。帯広市の委員会とその所管事項の例を図1に示す。

これらの委員会名と所管事項から、概念体系を作成する手順を以下に示す。

1. 名称末尾の「委員会」を削除
2. 名称末尾の「に関する事項」を削除
3. 名称末尾の「に属する事項」を削除

連絡先: 木村泰知, 小樽商科大学, 小樽市緑3丁目5-21,
kimura@res.otaru-uc.ac.jp

委員会名:総務文教委員会 調査事項 重要政策の企画及び総合計画に関する事項 財務に関する事項 広報及び広聴に関する事項 総合的な行政の推進に関する事項 ...

図1 帯広市の委員会とその所管事項の例

4. それぞれの上位概念に属する概念として「その他」を追加
5. 上位概念に並列な概念として「その他」を追加

上記の手順により、政治的カテゴリを作成し、平成 17 年度の小樽市の市議会会議録の第 1 回定例会および第 2 回定例会に対して、含まれていない政治的カテゴリが存在するか調査を行い、政治的カテゴリの修正を行った。その結果、「総務文教」、「厚生」、「産業経済」、「建築」、「その他」の 5 つの大カテゴリからなる 96 の政治的カテゴリを作成した。表 1 は、作成した政治的カテゴリの例である。

3. 会議録に対する政治的カテゴリの付与

本節の目的は、2. で作成した政治的カテゴリが、会議録に含まれる議員の発言にどの程度含まれているかを明らかにし、政治的カテゴリの観点から会議録や議員の特徴を分析することである。

まず、政治的カテゴリが含まれる発言の単位について説明する。本研究の目的からは、議員の活動または意見単位で政治的カテゴリを判断することが望ましいが、会議録において、活動や意見の単位で記述されているとは限らない。しかしながら、一般に、議題には議員の活動や意見が反映されていると考えられ、会議録ではある一つの議題に関する発言が一段落にまとめられる傾向にある。そこで、本分析では、段落単位で政治的カテゴリを判断することとした。

分析の対象は平成 19 年の小樽市市議会の会議録を対象として、1 年に 4 回開かれる定例会の会議録を分析することとした。平成 19 年の 4 回分の会議録は、総段落数 7,084 段落、文字数 1,034,664 文字、総単語数 651,643 単語、異なり単語数 10,344 単語である。分析作業は主観的判断になるため、2 名で同一の会議録を分析することで、可能な限り客観的な分析結果に近づけることとした。このような理由から、会議録の分析は、4 回の定例会に対して 8 人で分析を行った。表 2 に分析対象、段落数、分析者を示す。

会議録の分析は、平成 19 年度の小樽市の市議会会議録第 1 回から第 4 回までを対象としているが、平成 19 年 4 月に市議会議員選挙が行われたため、第 1 回定例会(3 月)の議員と第 2~4 回の議員が異なる。そこで、2 期に渡って市議会議員の職に就いている議員を対象とすることとした。また、市長、議員以外の発言者、発言していない議員については対象外とした。また、議員名の記述に誤りがあるものを訂正した結果、対象議員は 18 名となった。

表 3 は、政治的カテゴリと議員のクロス表であり、対象議員 18 名に対する発言の多い上位 13 カテゴリである。

ただし、このカテゴリ数は、分析者が一つでも注釈付けしたカテゴリの合計であり、2 名の分析者による重複したカテゴリは 1 回として数えていえる。これは、2 名による同一の注釈と 1 名だけによる注釈が存在するためである。

表 1 政治的カテゴリの例

カテゴリ番号	大カテゴリ	中カテゴリ	小カテゴリ
1000	総務文教		
1010		財務	
1011			地方税
1012			予算
1013			地方債
1020		総合的な行政の推進	
1021			条例
1022			高齢化対策

表 2 会議録の分析対象

分析対象	段落数	分析者
定例会第 1 回	1,747	男性 1 名 女性 1 名
定例会第 2 回	1,781	男性 1 名 女性 1 名
定例会第 3 回	1,711	男性 1 名 女性 1 名
定例会第 4 回	1,849	女性 2 名
合計	7,084	男性 3 名, 女性 5 名

表 3 の内容をみると、「財務」に関する内容が第 1 位となっており、全ての議員に対して発言数が多いことがわかる。「財務」に関する内容は議員に関係なく議論されていると考えられる。一方、2 位の「病院事業」に関しては、地域特有の議題である。なぜなら、一般的な病院に関する議題ではなく、小樽市立病院の移転に限定した内容となっているためである。このように発言数も多く、注目された議題の場合、賛否を考慮することも必要と考えられる。

次に、各議員の発言数の違いについて述べる。最も発言数が多かった議員 A は 829 回と最下位の議員 L の 86 回と比較して 10 倍近い差があった。勿論、所属する会派の影響などもあるため、発言数をもって単純に優劣を論じることはできない。しかしながら、カテゴリ単位で見た場合、議員ごとに発言内容の偏りが見られる。例えば、議員 A は「教育」や「学校」、「住民活動」についての発言が比較的多く、他の議員よりも力を入れていると考えられる。また、発言数が比較的小さい議員 G の「職員」や議員 M の「観光」のように、カテゴリ単位で見た場合に多く発言していることを考慮することで、議員が重要視する政治的カテゴリを特徴づけることができると考えられる。このような特徴を活用することで、「観光客が来なくて困っている」住民に対しては、議員 A よりも議員 M, G, F に関する意見や活動情報を提示するといったことが可能になると考えられる。

また、議員単位で見た場合、議員内の発言数の多いカテゴリに注目することで、議員を特徴づけることが可能と考えられる。我々は、平成 19 年 8 月に小樽市の市議会議員に対して、各議員の重視した活動に関するアンケート調査を実施している[1][2]。このアンケートは、重視した活動をキーワードあるいは、キーワードで 10 個考えてもらい、優先順位を付けして回答してもらったものである。対象議員は平成 11 年および平成 15 年に当選した市議会議員 37 名を対象としたものであり、有効回答数は 15 件であった。このアンケート結果を利用して、会議録の分析結果との比較を行う。表 3 の 18 名の議員の中で回答した議員は 7 名であり、その 7 名の議員が回答した 10 件の活動を正解として、今回の分析結果と比較することとした。議員の回答は自由記述であり、カテゴリに対応していないため、適切なカテゴリが含まれているかを人手で確認した。

表3 会議録を対象とした各議員の政治的カテゴリ数

順位	カテゴリ名	各議員の発言数 (総発言数の多い議員順に並べている)																		合計
		A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	
1	財務	94	122	44	73	40	43	30	18	17	19	20	32	10	19	22	17	9	4	633
2	病院事業	26	63	25	59	19	21	3	19	13	2	8	7	1	15			5	6	292
3	教育	39	2	22	19	18	22	4	1		23	9	4	1	10	25	15	14	8	236
4	総合的な 行政の推進	23	20	24	20	19	24	22	24	9	7	2		3	8	4	10	3	3	225
5	学校	53	14	20	19	12	19	5	1		19	5	4		3	14	10	7	5	210
6	医療	24	43	2	38	21	17	2	3	13	7	9	5		14	3		6		207
7	施設	25	27	22	6	24	26	10	9	7	3	6	7	17	2				4	195
8	職員	10	18	17	14	8	18	20	19	2	10	6			2	6	9	4	2	165
9	予算	22	33	20	12	5	9	12	5	10	12	2			3	4	2	7		158
10	住民活動	23	15	13	11	20	2	12	4	5	12	9	2	2	11	4	1	2		148
11	地域社会	16	10	23	19	8	7	9	22	2	6	5		4	1	1	5	6		144
12	観光	14	1	10	8	1	15	19			3	1	6	31	1	5	1	9	7	132
13	建築	21	24	16	14	8	16	3	3	6		2	1	4			3	5	3	129
～ 途中省略 ～																				
	総計	829	803	605	541	527	464	326	238	213	182	175	159	143	140	130	130	111	86	5802

表4 議員 M のアンケート結果と分析結果の比較

順位	アンケート結果	分析結果	カテゴリ数
1	水と緑を歴史を生かした <u>まちづくり</u>	観光	31
2	土産物 <u>観光</u> から生活 <u>観光</u> へ産業としての <u>観光</u> の確立	施設	17
3	観光の広域ネットワーク化と新たな交流 <u>観光</u> 拠点の創出	まちづくり	12
4	歩行者中心の交通ネットワークの形成推進	財務	10
5	都市修景による新しい <u>まちづくり</u>	経済	8

表5 ブログを対象とした各ブログサイトの政治的カテゴリ数

順位	カテゴリ名	各サイトの発言数 (注釈数の多いサイト順に並べている)																	
		ア	イ	ウ	エ	オ	カ	キ	ク	ケ	コ	サ	シ	ス	セ	ソ	タ	チ	ツ
1	情報	99	83	40	32	41	23	55	18	22	5		69	89	49	26	64	13	6
2	観光	41	63	48	2	27	6	63	29	4	25	92	19	20	33	13	29	2	42
3	地域社会	68	89	33	47	57	7	37	62	8	6	41	12	6	26	19	20	17	16
4	文化	45	21	33	20	33	28	56	32	23	11	2	6	5	14	3	5	3	6
5	施設	67	20	29	30	31	39	84	36	24	11			1	1	4	1	2	13
6	まちづくり	48	11	29	18	27	3	5	6	1	4	74	5	7	5	2	2	6	3
7	職員	43	12	35	52	11	16	50	1	5	3							3	
8	商業	26	7	15	1	13	10	2	2	4	4		53	1	3	6	1	1	
9	道路	24	24	20	3	22	28	12	10	20	30					2			2
10	経済	42	1	21	24	7	46	2	1	24	1		1						2
～ 途中省略 ～																			
		1073	856	738	720	646	642	594	436	385	278	270	258	204	199	189	186	183	179

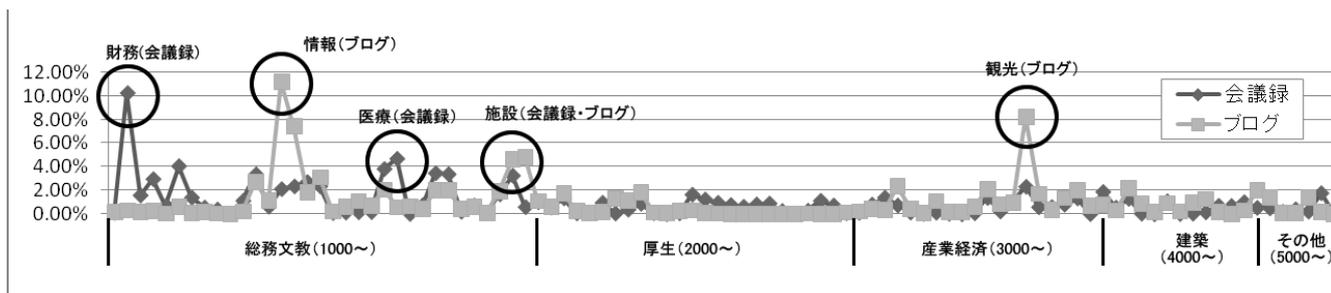


図2 会議録とブログを対象とした政治的カテゴリの頻度分布

その結果、74.2% (=52 の適切なカテゴリ ÷ (7 議員 × 10 回答)) が対応していた。議員 M に対してアンケート結果と分析結果を比較した例を表 4 に示す。議員 M の回答は、表 4 から見てわかるように、アンケート結果に「観光」と「まちづくり」に関する記述が多く、力を注いでいたことがわかる。また、人手による分析結果も「観光」と「まちづくり」が上位にあり、議員内の発言数が議員の活動をあらわしているといえる。

4. ブログに対する政治的カテゴリの付与

前節で述べた会議録の分析対象と合わせるために、小樽住民のブログを対象にする。ブログ収集に関する予備調査の結果、ブログの内容から地域を特定することは単純にキーワードを利用する程度では、収集精度が低いことが確認されたため、人手によりブログの内容を判断し、URL を特定した後に、自動収集をすることとした。その結果、小樽に関するブログを 40 件見つけることができ、それらの各ブログに含まれる記事を全て自動で収集した。ブログ間のバランスを保つために、各ブログから最大 100 記事を抽出し、合計で 2,581 件を分析対象とした。

次に、分析方法について説明する。分析方法は会議録と同様の方法で進める。ここでは、会議録の分析方法と異なる個所だけ説明する。分析単位は、会議録のように段落単位ではなく、ブログの記事(1日の投稿内容)とした。また、ブログ記事の内容は政治的課題と関連が薄い傾向にあることが予備調査から確認されていたため、半ばこじつけるように連想される政治的カテゴリを判断するように指示した。分析者は会議録と同じく大学生 4 名で、2 名ずつ同一のブログ記事を判断するようにしたため、1 名の分析者は約 1,300 記事を担当した。

表 5 は、各ブログサイトに対する政治的カテゴリの頻度である。ブログは 40 サイトを分析対象としたため、カテゴリ数の多い上位 18 件のみを表示する。ブログに付与された政治的カテゴリは、96 カテゴリ中 94 カテゴリであり、「上下水道」、「年金保険」の 2 カテゴリのみ付与されなかった。この結果から、今回利用したカテゴリが、ブログに対しても付与できることを確認した。

ブログ記事に対して、「情報」が最も多く付与されたカテゴリであった。この原因としては、「情報」というカテゴリが曖昧であり、ブログに記述されている内容には、何かしら情報に関する記述があり、利用する場合は、何に関する情報なのかを明確にする必要がある。

また、ブログの政治的カテゴリが「観光」として注釈されたサイトの上位 3 件は、(サ)、(イ)、(キ)がある。(ア)のブログサイトは、冬のイベント「雪明かりの路」オフィシャルブログであり、(キ)についてはホテルが毎日ホテルに関する情報を発信していた。これら 2 つのブログは個人のブログではなく、本研究の市民とのマッチングを目的としている点に適していない。しかしながら、(イ)については個人のブログであり、小樽を散策する内容となっており、地域に関する情報、観光情報などが記述されていた。この結果

から、個人のブログとそれ以外のブログを区別する必要があると考えられる。

5. 会議録とブログのマッチングへ向けた検討

会議録とブログのカテゴリの政治的カテゴリの割合を図 2 に示す。この会議録データは 3. で例にした 18 人以外の議員も含めた割合である。図 2 からわかるように、会議録の「財務」と「医療」、ブログの「情報」と「観光」がそれぞれ突出していることがわかる。しかしながら、「施設」のように会議録とブログに同程度の割合で出現するカテゴリがあること、今回設定したカテゴリが会議録においては全て付与されており、ブログにおいても 2 カテゴリを除いて付与されたことから、マッチングできる可能性が確認できた。

また、「財務」と「情報」については、マッチングの際に曖昧性の観点から問題と考えられる。しかしながら、政治に関する内容が少ないブログを大まかに分類できる利点がある。そして、曖昧解消としては、単独のカテゴリによるマッチングを考えるのではなく、他のカテゴリも考慮することで、マッチングすることが可能と考えられる。

6. 結論

本稿では、マッチングへ向けて、市議会会議録およびブログに対する政治的カテゴリを付与する分析を行った。その結果、議員を特徴づけるカテゴリの抽出方法については、議員内の発言数が多いカテゴリを利用することが有効であることが確認された。また、ブログ内にも政治に関するカテゴリを付与できることが確認できた。

謝辞

本研究の一部は総務省 SCOPE 補助金(No.082301004)の支援により行われた。

参考文献

- [1] 木村泰知, 渋谷英潔, “ブログに潜在する政治的意見と議員活動とのマッチング手法”, 電子情報通信学会言語理解とコミュニケーション(NLC)研究会, pp.19-23, 2008.
- [2] 木村泰知, 渋谷英潔, 高丸圭一, “ブログを対象とした地域住民の政治的関心の分析”, FIT2008 情報科学技術フォーラム一般講演論文集 第 2 分冊, pp.149-150, 2008.
- [3] 長谷川大, 乙武北斗, 木村泰知, 渋谷英潔, 高丸圭一, 荒木健治, “市議会会議録を対象とした概念体系構築へ向けた分析”, 情報処理学会 研究報告(2008-NL-187), pp.23-28, 2008.