

地方政治情報システムのための 地方議会会議録への注釈付けタグセットの提案

大城 卓[†] 渡邊 裕斗[†] 渋谷 英潔[§] 木村 泰知[¶] 森 辰則[§]

[†]横浜国立大学 大学院 環境情報学部 [‡]横浜国立大学 工学部 [§]横浜国立大学 大学院 環境情報研究院
[¶]小樽商科大学 商学部 社会情報学科

E-mail: [†]oshiro@forest.eis.ynu.ac.jp, [‡]yuto0105@forest.eis.ynu.ac.jp,
[§]{shib,mori}@forest.eis.ynu.ac.jp, [¶]kimura@res.otaru-uc.ac.jp

1 はじめに

地方政治における議員やその活動に関する住民の認知度は、国政と比較して一般に低い傾向がある。マスメディアで取り上げられる政治情報は国政に関するものが多く、地方政治に関する情報は少ない。地方政治に関する情報源としては地方議会会議録があり、検索システムとして公開している自治体も多い。しかし、公開されている会議録は1回の会議に限っても膨大な量のテキストデータであり、その全てを読むには多大な労力がかかる。検索システムを利用すればキーワード検索で文を絞り込むことも可能だが、地方政治に関する知識がなければ適切なキーワードを入力することは困難であり、キーワードが含まれる文が必ずしも利用者が求める情報であるとも限らない。

そうした現状を受け、我々は地方政治に関する情報、特に地方議会会議録における議員の発言を中心とした政治情報システムの開発を行っている。このシステムでは、利用者は施策・事業のリストの中から関心のあるものを選択する。システムは、会議録中の、選択された施策・事業に関する意見を表示する。ここでさらに利用者は、選択した施策・事業に対する意見（賛成・反対）を選択する。この操作を何度か繰り返すことで、システムは利用者に近い考えを持つ議員を判断し、表示する。

システム実現のためには、会議録から施策・事業に関する意見抽出するとともに、利用者の選択した施策・事業および意見と、会議録中の施策・事業に関する意見の対応を判定する必要がある。この判定は、選択された施策・事業の文字列が含まれる文を会議録から探すだけでは実現することはできない。会議録中の、施策・事業に関する意見を表す表現が持つ構造を分析し、その構造を考慮した上で構成要素ごとに抽出を行い、構成要素ごとの対応を判定する必要がある。本研究では、会議録中の、施策・事業に関する意見の構造を考慮した注釈付けタグセットを定義し、そのタグセットに基づくタグの自動付与によって、会議録中の意見の自動抽出を目指す。

そこで、本稿では、システムの実現に適した会議録への注釈付けタグセットの提案を行い、そのタグセットの有効性を確認する。また、タグの自動抽出に向けた調査も行う。

2 関連研究

会議録を対象とした研究としては、以前より国会会議録を対象とした研究が行われてきた。川端ら[1]や山本ら[2]は、特徴的な表層表現を手がかりに国会会議録を対象とした自動要約を行っている。地方議会会議録を対象とした研究としては、木村ら[3]は、ウェブ上の情報から住民の関心に合わせた地方議会議員の情報を提示するシステムの開発を行っている。

主観的な情報の注釈付けとしては、Wiebeら[4]は、意見などの個人状態をニュース記事にフレーズレベルで注釈付けしている。また、評判情報に関する研究では、小林ら[5]をはじめとし、主観的評価の構成要素を細かく

く分類した上での注釈付きコーパスの作成が行われている。

また、森ら[6]は、統計量を構成する要素を細かく分類して定義することで、新聞記事からの統計量表現の自動抽出精度を向上させている。

これらの関連研究に対し、本研究では地方議会会議録を対象とし、会議録中の意見の自動抽出を目指す。そのために、会議録中の意見を構成要素に細かく分類して注釈付けるためのタグセットを提案する。

3 政治情報システムの一例

我々が開発を進めている政治情報システムのイメージを図1に示す。このシステムは、住民が、関心を持つ施策・事業に関する意見を会議録から容易に得られるようにすること、住民自身に近い考えを持つ議員を探し出すことを目的としている。

図1: 政治情報システムのイメージ

まず、利用者は自治体の主要な施策・事業の一覧(図1の左側)から興味・関心のあるものを選択する。システムは選択された施策・事業に対応する会議録中の意見を、賛成意見と反対意見に分類した上で表示する(図1の右側中央)。ここで利用者が施策・事業に対する賛成・反対を選択する(図1の右側上部)ことで、利用者に考えの近い議員を探し出し、順位を付けて表示する(図1の右側下部)。

我々は、このシステムにおける主要な施策・事業の一覧として、総合計画の情報を利用することを考えた。総合計画*とは、地方自治体における行財政運営の長期的な指針となる、市政運営の基本となる計画であり、その中では、自治体の将来目標とそれを実現するために必要な施策・事業が体系的に明らかにされている。この施策・事業の体系に、対応する会議録中の意見を分類し、

*宇都宮市の総合計画 <http://www.city.utsunomiya.tochigi.jp/machizukuri/sougoukeikaku/index.html>

さらにその意見を賛成・反対に分類することで、システムの実現を目指す。

なお、総合計画では具体的な取り組みを事業、事業を一定のグループにまとめたものを施策としているが、どこまでを具体的な取り組みとするかは自治体によって異なる。また、現在開発を進めているシステムでは施策・事業が体系化されていることが重要であり、施策と事業を明確に分ける必要はない。そのため、本稿では施策と事業を特に区別せず、どちらも事業と呼ぶこととする。

4 事業に関する意見が持つ構造

3章で説明したシステムを実現するためには、会議録から事業に関する意見を抽出するとともに、利用者の選択した事業およびその事業に対する意見（賛成・反対）と、会議録中の事業に関する意見の対応を判定する必要がある。そのために本研究では、会議録中の、事業に関する意見を表す表現が持つ構造を考慮した上で構成要素ごとに抽出を行い、構成要素ごとの対応を判定することを目指す。本章では、会議録中の、事業に関する意見を表す表現が持つ構造の分析を行う。

会議録中に存在する、事業に関する意見を含む例文を図2に示す。

1. (a) パーク・アンド・ライドの
(b) 実施が必要です。
2. (c) 地域経済の (d) 活性化に つ き ま し て、
(e) 検討を (f) 進めてまいります。
3. (g) 高齢者の (h) 外出(i) 支援を (j) 充実させ、
4. (k) 予防ワクチンの (l) 接種に対する (m) 助成
を早期に (n) 実施すべきであると考えます。
5. 短期保険証の (o) 期間は (p) 3カ月とすべきで
はないでしょうか。
6. この事業の継続には (q) 賛同できません。
7. (r) 宝木町2丁目の新川沿いに調整池用地を購
入するのは大いに (s) 評価するところであります。

図2: 事業に関する意見を含む例文

事業に関する意見は、例文(1)の下線部(a)のように事業を表す表現と、下線部(b)のように、発言者（議員、自治体の首長および部長、議長など）の主張や考えを表す表現に分けられる。まず、事業を表す表現について考える。例文(1)では、事業は下線部(a)の名詞1語で表されているが、例文(2)では、下線部(c)と下線部(d)の2語を含むより大きな名詞句で1つの事業が表現されている。ここで、下線部(c)と下線部(d)は項と述語の関係にある。例文(3)では、下線部(g)、下線部(h)、下線部(i)を含む名詞句で1つの事業が表現され、下線部(g)と下線部(h)を含む名詞句と下線部(i)が項と述語の関係にある。例文(4)では、下線部(k)、下線部(l)、下線部(l)を含む名詞句で1つの事業が表現され、下線部(k)と下線部(l)を含む名詞句と下線部(l)が項と述語の関係にある。例文(2)、(3)、(4)でそれぞれ、事業を表す表現の述語にあたる部分（下線部(d)、下線部(i)、下線部(l)）に注目すると、どの表現も動作主が行政機関であることがわかる。また、下線部(d)、下線部(i)、下線部(l)はどれも、ある対象物に意図を

持って変化を加えることを表す表現である。このように、事業とは、行政機関がある対象物に何らかの変化を加えること、あるいは「対象物に対する変化」の集合であると捉えることができる。行政が加える変化の内容に注目すると、例文(3)では下線部(h)の“支援”，例文(4)では下線部(l)の“助成”である。変化を加える対象物はそれぞれ2語を含む名詞句であるが、例文(3)では下線部(f)と下線部(g)は動作主と述語の関係にあるのに対し、例文(4)では下線部(j)と下線部(k)は対象物と述語の関係にある。このように、変化を加える対象物にあたる部分は名詞1語とは限らず、動作主・対象物・述語といった述語項構造に相当する表現で表されることもある。また、事業を構成する構成要素は他にも、例文(5)の下線部(o)のように変化を加える対象物のある側面を表す表現や、下線部(p)のように変化後の状態を表す表現が現れることもある。

次に、発言者の主張や考えを表す表現について考える。まず、例文(2)の下線部(e)や例文(3)の下線部(i)のように、事業の「方向性」を表す表現が現れることがある。そして、既に記されている事業の方向性について、例文(1)の下線部(b)や例文(2)の下線部(f)のように、実際にその方向で実行するか否かを表す「態度」の表現が現れることがある。発言者が行政機関側であれば実行する・しないという表現になるが、発言者が議員の場合は、議員は実際に実行する権限を持たないので、例文(1)の下線部(b)や例文(4)の下線部(n)のように実行を要求する表現になる。さらに、事業の方向性に対する態度について、例文(6)の下線部(q)や例文(7)の下線部(s)のように賛成・反対、良い・悪いといった「評価」が続くこともある。

これらの構成要素には、それぞれに例文(7)の下線部(r)のような修飾部が存在する場合がある。下線部(r)は、修飾している構成要素に限定を加える表現である。事業を表す表現が修飾部によって限定されている場合は、修飾部の有無や、その内容によって発言者の主張や考えが異なる場合もあるため、修飾部も抽出することができれば有用な情報であると考えられる。

5 事業に関する意見を注釈付けるためのタグセットの提案

5.1 タグセットの構成要素

4章の分析に基づき、事業に関する意見を大きく2つの部分に分けた上で、それぞれの構成要素を分類して抽出することを目的として、会議録に注釈付けを行うためのタグセットを提案する。提案するタグセットを以下に示す。

● 事業を構成する構成要素

- <subjectA>: 主題（動作主）
取り上げられているもの・ことのうち、動作主にあたる部分
- <subjectO>: 主題（対象物）
取り上げられているもの・ことのうち、対象物にあたる部分
- <subjectP>: 主題（述語）
取り上げられているもの・ことのうち、述語にあたる部分
- <aspect>: 側面
主題のある側面
- <goal>: 到達点
変化後の状態
- <change>: 変化
行政が主題に加える変化の内容

<enterprise>：事業
主題・側面・到達点・変化の組が1つの（複
合）名詞で表現されているもの

- 発言者の主張や考えを構成する構成要素

<orientation>：方向性
事業に対するアクション

<attitude>：態度
事業・方向性の組に対する意思表示（実行す
る・しない）

<evaluation>：評価
事業・方向性・態度の組に対する主観的見解
（賛成・反対、良い・悪い）

- その他

<modify>：修飾部
各構成要素の修飾部

各タグにはそれぞれ一意に決まる識別子を id 属性の
値として付与する．1つの事業を構成する構成要素には
同一の組番号を値として持つ tuple 属性を付与する．1
つの「発言者の主張や考え」を構成する構成要素には同
一の組番号を値として持つ tuple 属性と、肯定・否定・
その他のいずれに該当するかを値として持つ polarity 属
性を付与する．修飾部には、修飾している構成要素の識
別子を target 属性に付与する．各構成要素間の関係を
図3に示す．各構成要素は省略される場合がある．

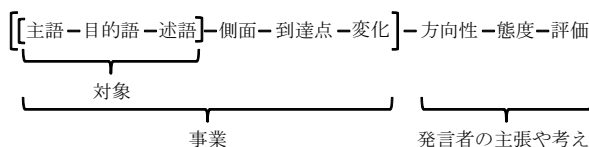


図 3: 構成要素間関係

5.2 コーパスの試作

前述のタグセットに基づき、注釈付きコーパスの試作
を行った．注釈付けには平成 22 年宇都宮市議会第 1 回
定例会の会議録 4056 文を用いた．この議会には 1 日目
から 6 日目までの会議録が存在している．その内容は全
て一様というわけではなく、1 日目は、市政運営の基本
方針と平成 22 年度の主要施策の概要について市長から
の説明、2 日目から 5 日目は、質疑と一般質問、6 日
目は、議案に対する賛成・反対討論および採決が主な内
容となっている．作成した注釈付けコーパスの一部を図
4に、4056 文のコーパスに注釈付けされた各構成要素の
タグの総数を表 1に示す．

今後とも、<subjectO id="so917_1" tuple=
"t917_1"> ごみ </subjectO> の <subjectP id=
"sp917_1" tuple="t917_1"> 分別 </subjectP>
の <change id="c917_1" tuple="t917_1"> 徹底
</change> が <attitude id="at917_1" polar-
ity="p" tuple="t917_1"> 必要 </attitude> と
思われますが、今後の対策をお伺いいたします。

図 4: タグ付与の例

6 タグセットの活用事例

住民がある事業に関する意見を会議録から検索した
いと考えた場合、会議録検索システムが存在する自治体

表 1: 注釈付けされたタグ数

	総数	異なり数
subjectA	14	9
subjectO	920	633
subjectP	154	82
aspect	140	94
goal	35	26
change	996	413
enterprise	376	226
orientation	316	129
attitude	832	228
evaluation	69	61
modify	1129	834

であれば、現在でもキーワード検索による文の絞り込み
は可能である．しかし、キーワードを含む文に必ずキー
ワードに関する意見が含まれているとは限らず、検索
結果には純粋な質問や客観的事実を述べている文、キー
ワードに関連していない意見を含む文なども存在してい
る．一方で、会議録に 5 章で定義したタグが付与されて
いれば、タグが付与されている箇所のみを検索対象とす
ることで、検索結果をキーワードに関する意見を含む文
のみに絞り込むことができる．

そこで、タグセットの有効性を確認する一例として、
会議録に多く現れるキーワードを検索クエリとし、全文
検索を行った結果とタグが付与されている箇所のみを検
索対象文字列として検索を行った結果を比較し、どの程
度絞り込みができていたかを調査する．

5.2 節のコーパス全体を対象として、2 つの検索方法
で検索を行った場合に、それぞれの検索結果に含まれる
文の数を表 2 に示す．今回選択したキーワードでは、タ
グが付与されている箇所のみを検索対象とすることで、
検索結果の文数を 2 割から 3 割程度まで絞り込むことが
できた．よって、事業に関する意見を会議録から検索す
ることを目的とした場合には、会議録にタグを付与する
ことによって、全文検索の場合と比較して検索効率を大
幅に向上させることができると考えられる．

表 2: 検索結果

	全文	タグ箇所のみ
中心市街地	70	16
高齢者	57	13
雇用	55	11
自転車	34	10
公共交通	30	9
LRT	26	6
防犯	25	5
生涯学習	22	7

7 事業に関する意見の自動抽出に向けた調査

5 章で定義したタグセットの各構成要素について、標
準的な情報抽出手法によってどの程度の精度で抽出が可
能か調査する．森ら [6] は、文字を単位とする系列ラベ
リングに基づく、標準的なチャンキング手法を用いて精
度向上を実現している．そこで、森ら [6] の手法と同等
のものを用いて、構成要素の自動抽出実験を行う．ここ
では、modify については、修飾している構成要素との
係り受け構造に基づいて抽出すべきものであると考え、
チャンキングによる自動抽出は行わない．

7.1 実験の設定

チャンキングには、SVM に基づく汎用チャンカー
である YamCha[7] を使用し、エンコーディング法には
IOB2 法 [8] を利用し、チャンキングの解析方向は左向

表 3: 構成要素抽出の適合率, 再現率, F 値

タグ名	限定なし			学習データを限定			学習・テストデータを限定		
	Pre	Rec	F	Pre	Rec	F	Pre	Rec	F
subjectA	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
subjectO	16.4	7.3	10.1	9.9	19.9	13.3	32.7	19.9	24.7
subjectP	13.9	2.1	3.7	14.0	8.4	10.5	26.0	8.4	12.7
aspect	25.1	8.1	12.2	12.2	17.4	14.3	34.0	17.4	23.0
goal	36.7	6.7	11.3	36.3	18.4	24.4	63.3	18.4	28.5
change	25.9	11.6	16.0	17.1	29.1	21.5	47.9	29.1	36.2
enterprise	19.7	10.1	13.4	14.8	21.1	17.4	37.4	21.1	27.0
orientation	46.9	26.8	34.1	35.4	46.0	40.0	66.5	46.0	54.4
attitude	48.5	23.7	31.9	36.8	46.7	41.2	74.5	46.7	57.4
evaluation	31.5	10.1	15.3	17.6	26.5	21.1	61.4	26.5	37.0

きとし、文脈長は対象文字の前後2文字ずつ、計5文字とした。素性には、文字自身、文字種、品詞、形態素、文節内素性、複合名詞主辞素性を用いた。文節内素性は、文節内に固有表現が存在すれば、最も先頭に近い固有表現の品詞細分類を、固有表現がなければ文節の先頭の形態素を用いるものである。複合名詞主辞素性は、連続する名詞が存在する場合、連続する名詞の最後の名詞を素性とするものである。

実験データには5.2節で述べたコーパスを用いた。コーパス全体を5個の部分コーパスに分割して5分割差検定を行い、適合率、再現率、F値で評価した。

5.2節で述べたコーパスでは、事業に関する意見に対して注釈付けを行うという方針であったために、事業を構成する構成要素のうち、発言者の主張や考えを構成する構成要素と組になっている構成要素のみにタグが付与されており、事業を表す表現が単独で現れた場合にはタグを付与していない。そのため、同じように事業を表す表現であってもタグが付与されているものとないものがコーパス中に混在してしまっている。この注釈付けの方針が自動抽出に及ぼす影響を調査するため、学習データを構成要素のタグを含む文に限定した場合と、学習データおよびテストデータを構成要素のタグを含む文に限定した場合においても同様の実験を行った。

7.2 結果と考察

各タグの自動抽出における適合率、再現率、F値を表3に示す。

まず、学習データ・テストデータ共に限定していない場合の結果を見ると、どの構成要素も再現率が低く、適合率についても orientation と attitude が50%程度で、他の構成要素はさらに低い値になっている。学習データを構成要素のタグを含む文のみに限定した場合には、ほぼ全ての構成要素で再現率が向上するものの、その分適合率が低下し、F値についてもわずかに向上している程度である。学習データとテストデータを共に限定した場合には、適合率・再現率共に向上するが、それでもF値で50%を超えているのは orientation と attitude のみである。

この結果を見ると、現時点で比較的高い精度で抽出できている orientation と attitude については与える素性や学習器の検討等により高い精度での自動抽出が期待できるが、他の構成要素については標準的なチャンキング手法による自動抽出は難しいものと予想される。これらの構成要素は、会議録に人手で注釈付けを行う際にも係り受け構造を手がかりにタグを付与しているため、係り受け構造を利用した文節単位での自動抽出手法の導入が必要であると考えられる。また、コーパス内の事業を表す表現すべてに注釈付けを行うことで、精度向上を実現できる可能性がある。しかし、subject, aspect, goal, change といった構成要素が会議録中に単独で存在していた場合、それが事業の構成要素となり得るかは人手でも判断がつかないため、どちらにせよ全ての構成要素の自動抽出を実現するためには係り受け構造を

利用した手法が必要であると考えられる。subjectA, goal, evaluation については表1からもわかる通り、会議録中での出現頻度が低く、それが低い精度の原因の一つであると考えられるため、学習データを増やして実験を行い、検証する必要がある。

8 おわりに

本稿では、地方議会会議録における議員の発言を中心とした政治情報システム開発のための注釈付けタグセットを提案し、その活用事例を示すとともに、タグの自動抽出の可能性についても調査した。

今後の課題としては、係り受け構造を利用した文節単位の構成要素抽出手法の検討がある。また、3章のシステムを実現するためには、抽出した意見を対応する施策・事業に自動分類する手法の検討や、orientation, attitude, evaluation の持つ極性の自動判定手法の検討が必要である。さらに、今回作成し、調査に使用したコーパスはある自治体における1回の定例会の会議録のみであり、臨時会や各種委員会といった別の種類の議会や、異なる年、自治体の会議録においてもそれらの手法が有効であるか確認する必要がある。

謝辞

本研究の一部は、科学研究費補助金(No.22300086)の助成を受けたものである。

参考文献

- [1] 川端正法, 山本和英. 話題の継続に着目した国会会議録要約. 言語処理学会第13回年次大会, pp. 696-699, 2007.
- [2] 山本和英, 安達康昭. 国会会議録を対象とする話し言葉要約. 自然言語処理, Vol. 12, No. 1, pp. 51-78, 2005.
- [3] 木村泰知, 洪木英潔, 高丸圭一, 武北北斗, 小林哲郎, 森辰則. 地方議員マッチングシステムにおける能動的質問のための質問生成手法. 人工知能学会論文誌, Vol. 26, No. 5, pp. 580-593, 2011.
- [4] Janyce Wiebe, Theresa Wilson, and Claire Cardie. Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation*, Vol. 39, No. 2-3, pp. 165-210, 2005.
- [5] 小林のぞみ, 乾健太郎, 松本裕治. 意見情報の抽出/構造化のタスク使用に関する考察. 自然言語処理研究会報告 2006-NL-171, 情報処理学会, 2006.
- [6] 森辰則, 藤岡篤史, 村田一郎. 動向情報編纂のためのテキストからの統計量表現の自動抽出. 人工知能学会論文誌, Vol. 23, No. 5, pp. 310-318, 2008.
- [7] Taku Kudo and Yuji Matsumoto. Chunking with support vector machines. *NAA CL*, pp. 1-8, 2001.
- [8] Erik F and Tjong Kim Sang. Noun phrase recognition by system combination. In *Proceedings of ANLP-NAA CL*, pp. 50-55, 2000.