Creating Japanese Political Corpus from Local Assembly Minutes of 47 Prefectures

Yasutomo Kimura
Otaru University of Commerce
`kimura@res.otaru-uc.ac.jp`

Keiichi Takamaru
Utsunomiya Kyowa University
`takamaru@kyowa-u.ac.jp`

Takuma Tanaka
Toyohashi University of Technology
`t-tanaka@la.cs.tut.ac.jp`

Akio Kobayashi
Toyohashi University of Technology
`a-kobayashi@cs.tut.ac.jp`

Hiroki Sakaji
Seikei University
`hiroki_sakaji@st.seikei.ac.jp`

Yuzu Uchida
Hokkai-Gakuen University
`yuzu@hgu.jp`

Hokuto Ototake
Fukuoka University
`ototake@fukuoka-u.ac.jp`

Shigeru Masuyama
Toyohashi University of Technology
`masuyama@tut.jp`

Abstract

This paper describes a Japanese political corpus created for interdisciplinary political research. The corpus contains the local assembly minutes of 47 prefectures from April 2011 to March 2015. This four-year period coincides with the term of office for assembly members in most autonomies. We analyze statistical data, such as the number of speakers, characters, and words, to clarify the characteristics of local assembly minutes. In addition, we identify problems associated with the different web services used by the autonomies to make the minutes available to the public.

1 Introduction

Many local autonomies in Japan provide access to various political documents on their websites. Such documents include basic urban development plans, local assembly minutes, and ordinances. The information obtained through the Internet can be used to compare the autonomies and identify the characteristics of individual autonomies. Local assembly minutes are an especially important source of such characteristics because they include various representatives' positions on policies enforced by the autonomy. Some studies that compare local assembly minutes have been conducted by political scientists (Tadashi Masuda, 2012). However, some issues arise with the analysis of local assembly minutes that should be addressed. One such issue is the different ways used to release the minutes to the public. There are 47 prefectures and several cities, towns, and villages in Japan, and local assembly minutes are made available in a variety of ways. Therefore, collecting local assembly minutes and unifying the format of the collected data for analysis on a national level is cost inefficient. In this paper, we attempt to create a corpus of local assembly minutes in Japan. This corpus realizes some research findings from NLP and sociolinguistics. Our objective is to develop a corpus that can be used for a broader range of interdisciplinary research.

We collected opinions from political scientists and economists to identify their research requirements with regard to local assembly minutes. It became clear that the scope of minutes must be controlled to conduct a strict comparative analysis. Specifically, the following requirements were identified as important: i) the periods must be identical and ii) autonomies must be classified as prefectures and other small regions such as cities, towns, and villages. For the first requirement, we collected minutes from assemblies held between April 2011 and March 2015. This four-year period is consistent with the term

Proceedings of the 12th Workshop on Asian Language Resources,
pages 78–85, Osaka, Japan, December 12 2016.

of office for assembly members' in most autonomies. To satisfy the second requirement, we collected assembly minutes from all 47 prefectures.

The objectives of this study are as follows: i) collect local assembly minutes (each of which have different formats) from all 47 prefectural assemblies, ii) obtain fundamental statistical data from the collected text data, and iii) extract political keywords using term frequency-inverse document frequency (tf-idf) weighting for a preliminary comparative analysis.

2 Collecting local assembly minutes

2.1 Outline

In this study, our primary purpose was to collect minutes of assemblies between April 2011 and March 2015 from all 47 prefectures. In the assembly minutes, all utterances are transcribed and recorded. Thus the assembly minutes will reveal policy planning and consensus building processes. As a result, assembly minutes are valuable documents for political analysis.

Local autonomies are independent, and the methods used to release assembly minutes differ, thus making the collection of assembly minutes. For example, a customized web crawler will be required for each autonomy.

The variety of text formats presents another difficulty. We need to extract assembly members' utterance from crawled web documents. Utterance attribute information, such as the speaker's name and the meeting's name and date must also be extracted. Thus, the format must be considered for each autonomy.

In the next section, we describe the procedure used to extract assembly minutes and create the corpus.

2.2 Procedure

As shown in Table 1, there are four types of primary web services used to release assembly minutes, and 42 of the 47 prefectures adopt one of these four primary web services. We prepare a web crawler for the primary four web services. However, many local autonomies have partially customized these services. We use semi-automated crawling depending on how the local autonomies release such documents. The remaining five prefectures have developed unique services, and assemblies that use such unique systems are downloaded using semi-automated crawling.

Table 1: Main Web services for providing local assembly minutes.

Name of Web service		Number of prefectures
Discuss	created by NTT-AT corporation	18
DB-Search	created by Yamato_Sokki center corporation	16
VOICES	created by Futurein corporation	6
Sophia	created by Kobe_Sogo_Sokki corporation	2
Others	Used by Iwate, Akita, Shizuoka, Wakayama and Okinawa	5

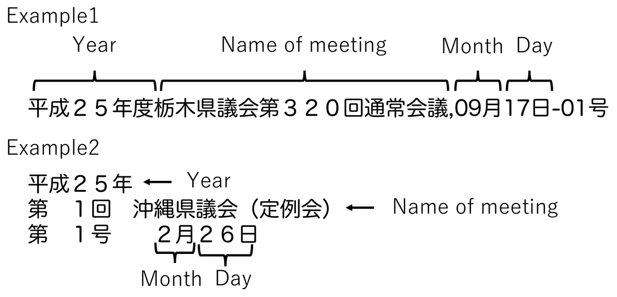
Table 2 lists the database items for the corpus. We automatically divided the raw into database items.

Below, we illustrate the method used to input text into the database after collecting the local assembly minutes. First, we designed a web crawler to automatically extract each value, such as "Year," "Month," "Day," and "Name of meeting," from a document. The following figure illustrates the extraction of "Name of meeting".

Table 2: Corpus database items.

Number	Name of item	Explanation
1	Reference number	Unique key for each utterance.
2	Name of prefectures	Name of autonomy.
3	Year	Year the meeting was held.
4	Month	Month the meeting was held.
5	Day	The day of the meeting.
6	Name of meeting	The name that includes the meeting type, volume and number.
7	Name of speaker	The name often includes a title such as “Mr.”, “Pres.” and “Chair.”
8	Utterance	The utterance is split by either “period” or “new line mark”.
9	Other	Other means outside the field of the utterance.

The description example of name of the meeting



Example1

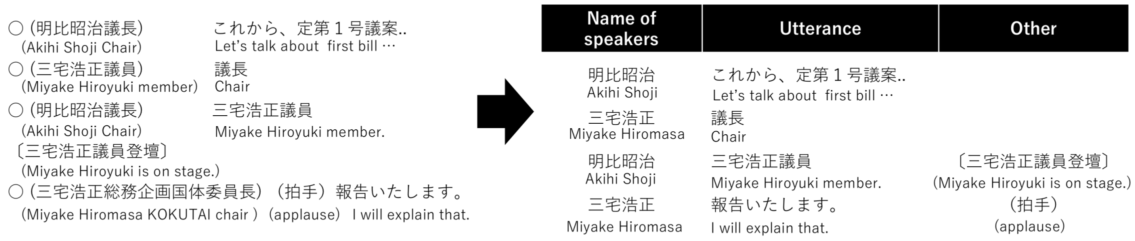
平成２５年度栃木県議会第３２０回通常会議,09月17日-01号

Example2

平成２５年
第　１回　沖縄県議会（定例会）
第　１号　　２月２６日

Second, we extract speakers’ names and their utterances from the body text of the minutes using a post-processing program. Then, the extracted values are registered to the database.

The example of body text of assembly minutes



○（明比昭治議長）　これから、定第１号議案..
（Akihi Shoji Chair）　Let’s talk about first bill …
○（三宅浩正議員）　議長
（Miyake Hiroyuki member）　Chair
○（明比昭治議長）　三宅浩正議員
（Akihi Shoji Chair）　Miyake Hiroyuki member.
〔三宅浩正議員登壇〕
（Miyake Hiroyuki is on stage.）
○（三宅浩正総務企画国体委員長）（拍手）報告いたします。
（Miyake Hiromasa KOKUTAI chair ）（applause） I will explain that.

Name of speakers	Utterance	Other
明比昭治 Akihi Shoji	これから、定第１号議案.. Let’s talk about first bill …	
三宅浩正 Miyake Hiromasa	議長 Chair	
明比昭治 Akihi Shoji	三宅浩正議員 Miyake Hiroyuki member.	〔三宅浩正議員登壇〕 （Miyake Hiroyuki is on stage.）
三宅浩正 Miyake Hiromasa	報告いたします。 I will explain that.	（拍手） （applause）

The body text of the assembly minutes does not intend to process by the program. The delimiters between speaker name and utterance differ among autonomies, with some of them having no delimiter. Some vary their delimiters arbitrarily. In such cases, we split the text manually.

3 Collecting minutes

In this section, we first clarify the characteristics of each prefecture by counting the words of the local assembly minutes. Then, we extract political keywords using the tf_idf weighting method for a preliminary study of comparative analysis.

3.1 Fundamental statistical data

Here, we show fundamental statistical data of the collected text data. Table 3 shows an overview of the minutes from the 47 prefectures. We explain two main categories in Table 3, i.e. “Prefectures” and “Speakers.” The “Prefectures” column includes the web service name, the numbers of speakers, characters, words, and different words. The “Speakers” column includes the maximum number of words, minimum number of words, and the average number of words for words and characters. Hence, we explain the definition of “words” for counting. The Japanese language does not use spaces to divide sentences into words. Therefore, we must write a Japanese sentence with some space between words or grammatical units prior to counting words. If we count either the number of words or different words, the result changes according to the word division method. The results of a word division method differ depending on the employed morphological analysis tool and dictionary. In Table 3, we use MeCab(Kudo et al., 2004) in combination with the IPAdic dictionary [1].

First, we focus on the minimum and maximum number of "speakers." Speakers includes local assembly members, a governor, and someone acting on behalf of the governor. In Table 3, the minimum and maximum number of "speakers" is 61 (Akita) and 1,091 (Saitama), respectively. In contrast, according to the "Ministry of Internal Affairs and Communications" website published in March 2012[2], the minimum and maximum number of "local assembly members" is 35 (Fukui and Tottori) and 125 (Tokyo), respectively. Although speakers include a governor and someone acting on behalf of the governor, the number of "speakers" was larger than that of assembly members because speaker's names were spelled in several different ways. For example, for the "Toda" local assembly member, four different representations of the name were found.

Second, we focus on minimum and maximum number of characters. In Table 3, the minimum and maximum number of characters is 2,191,589 (Yamagata) and 19,389,614 (Nagano), respectively. In prefectures having too many characters, the collected text includes non-utterances such as "symbols," "table of contents," "description of circumstance," and "supporting data." In particular, the size of "supporting data" is the largest one among non-utterances, and it is difficult to determine whether a sentence is an utterance.

Finally, we summarize the problems associated with creating a Japanese political corpus from local assembly minutes. Using the fundamental statistical data, we could not completely extract both "Name of speakers" and "Utterances." We faced two problems: i) a speaker's name was spelled in several different ways and ii) it was difficult to determine whether a sentence is an utterance.

3.2 Extracting political keywords using tf-idf

In this section, we attempt to detect specific political keywords from each prefecture. This is a pilot political comparative analysis study.

First, utterances in the minutes are divided into words by morphological analysis. We use MeCab (ipadic-NEologd) for this analysis. The target part of speech is a noun. We exclude "noun-dependent," "noun-number," "noun-pronoun," "noun-suffix" and "noun-proper noun-person's name" forms.

Then specific political words are calculated by applying tf-idf weighting. The TF is a term's occurrence within the minutes of each prefecture divided by the number of words in the minutes. The TF of term t for document d, where d denotes minutes, is expressed as follows. The minutes contain a significant number of words; therefore, we use a logarithmic value for tf.

$$tf(t,d) = \log \frac{count(t), t \in d}{|d|}$$

A term's DF is the number of documents (i.e., prefecture minutes) containing that term divided by the total number of documents. The DF of t in N documents is expressed as follows. In this case, N becomes 47.

$$df(t,N) = |\{d_i : t \in d_i,\ i = 1..N\}|$$

The multiplied value of TF and IDF is the score of the word.

Table 4 shows the top three words obtained using tf-idf weighting. We find many political keywords spoken in the assembly. For example "HAC," the first-ranked word in Hokkaido Prefecture, is a regional airline company whose largest shareholder is the prefecture. "Daisetsu Ribaa-netto," the third word in Iwate Prefecture, is an NPO that is known to commit crimes using subsidized money. "Shimane Japanese beef cattle," the second word in Shimane prefecture, is a bland of Japanese beef cattle which is promoted by this prefecture.

---

[1] https://osdn.jp/projects/ipadic/

[2] http://www.soumu.go.jp/main_content/000151136.pdf

Table 3: Overview of collecting minutes of 47 prefectures in Japan.

		Prefectures					Speakers					
		Name of service	Number of speakers	Number of characters	Number of words	Number of different words	Number of words			Number of characters		
	Prefecture						Max	Min	Ave	Max	Min	Ave
1	Hokkaido	VOICES	202	6,690,769	4,324,553	21,410	445	2	62	270	1	38
2	Aomori	DB-Search	230	6,075,988	3,845,805	24,085	529	1	63	320	1	38
3	Iwate	Other	115	5,188,754	3,284,943	21,281	883	1	62	517	1	37
4	Miyagi	discuss	106	6,492,978	4,153,106	26,876	1,116	1	52	695	1	32
5	Akita	Other	61	3,501,499	2,268,171	20,336	545	1	61	466	1	37
6	Yamagata	discuss	102	2,191,589	1,364,855	16,071	874	1	66	501	1	39
7	Fukushima	discuss	270	3,605,556	2,248,658	18,658	1,495	1	59	884	1	35
8	Ibaraki	DB-Search	387	4,378,426	2,730,223	21,772	783	1	63	437	1	37
9	Tochigi	VOICES	150	2,563,445	1,637,756	16,311	922	1	76	554	1	46
10	Gumma	VOICES	172	5,713,837	3,520,362	24,398	1,840	1	90	1,129	1	54
11	Saitama	discuss	1,091	6,280,996	4,125,088	26,586	431	1	43	263	1	26
12	Chiba	DB-Search	172	3,392,658	2,166,942	20,951	2,578	1	52	1,527	1	31
13	Tokyo	DB-Search	408	5,746,805	3,703,744	25,896	405	1	50	234	1	30
14	Kanagawa	discuss	156	5,896,670	3,682,485	23,896	575	3	58	349	1	34
15	Niigata	discuss	900	15,885,612	9,839,042	33,855	692	1	59	424	1	35
16	Toyama	DB-Search	124	4,694,955	2,941,880	23,252	813	1	63	482	1	38
17	Ishikawa	VOICES	158	4,413,772	2,767,061	22,049	916	3	111	536	2	67
18	Fukui	DB-Search	149	4,568,260	2,954,643	24,135	590	1	52	345	1	32
19	Yamanashi	DB-Search	164	4,274,363	2,823,520	19,892	499	1	54	285	1	34
20	Nagano	VOICES	524	19,389,614	12,092,538	36,162	1,691	1	94	1,004	1	57
21	Gifu	discuss	517	6,422,363	4,032,530	26,243	653	1	60	374	1	36
22	Shizuoka	Other	239	5,376,236	3,353,750	25,556	689	1	59	530	1	35
23	Aichi	DB-Search	304	5,881,919	3,685,199	24,982	530	1	67	304	1	40
24	Mie	discuss	115	4,979,765	3,074,270	24,639	642	1	66	371	1	39
25	Shiga	VOICES	249	8,626,218	5,417,961	28,167	1,475	1	93	900	1	56
26	Kyoto	DB-Search	765	14,714,871	9,094,335	33,386	981	1	72	577	1	42
27	Osaka	discuss	536	17,318,822	10,794,849	32,611	572	1	60	347	1	35
28	Hyogo	Sophia	154	3,892,396	2,433,087	22,687	688	2	90	435	1	54
29	Nara	discuss	114	4,134,566	2,596,260	22,266	474	1	58	300	1	34
30	Wakayama	Other	102	3,427,751	2,166,894	24,731	1,212	3	54	655	2	32
31	Tottori	DB-Search	168	10,844,070	6,726,931	35,631	994	1	63	577	1	37
32	Shimane	DB-Search	122	6,010,462	3,785,152	25,948	599	1	56	353	1	34
33	Okayama	discuss	103	6,296,556	3,962,654	26,490	752	1	57	421	1	34
34	Hiroshima	DB-Search	170	3,357,629	2,125,208	18,831	580	1	62	339	1	37
35	Yamaguchi	discuss	99	4,932,992	3,133,703	23,063	559	1	60	333	1	36
36	Tokushima	discuss	86	3,812,198	2,498,206	21,113	550	1	53	319	1	33
37	Kagawa	DB-Search	388	8,752,886	5,592,540	29,344	527	1	67	326	1	41
38	Ehime	Sophia	203	4,198,966	2,593,711	22,947	798	2	90	455	1	54
39	Kochi	discuss	92	5,879,641	3,641,928	24,618	411	1	67	241	1	39
40	Fukuoka	DB-Search	177	4,948,309	3,137,731	22,890	391	1	58	246	1	35
41	Saga	DB-Search	124	5,740,329	3,680,483	22,066	487	1	59	328	1	36
42	Nagasaki	discuss	676	12,806,907	8,002,185	31,101	773	1	55	441	1	32
43	Kumamoto	discuss	116	4,700,096	2,965,285	25,752	760	1	57	461	1	34
44	Oita	discuss	183	4,595,154	2,899,446	22,475	574	1	48	327	1	29
45	Miyazaki	discuss	106	6,471,745	4,027,718	27,998	393	1	62	244	1	37
46	Kagoshima	DB-Search	151	7,266,842	4,646,583	26,674	581	1	64	349	1	39
47	Okinawa	Other	153	7,553,407	4,702,333	26,855	926	1	50	521	1	29
Total	—	—	11,853	303,889,642	191,246,307	—	—	—	—	—	—	—

There are many compound words in Japanese political words. For example “zaisei-kaikaku (fiscal reform)” comprises two nouns, i.e., “zaisei (fiscal)” and “kaikaku(reform).” Therefore, we combine continuous nouns and attempt to handle them as a single compound word. Table 5 shows the top three compound words obtained by tf-idf weighting. We find many titles of the officer, i.e. “Mr.” and “Prefectural Assembly”. The title of an officer is typically derived from political agendas. For example, the third words in Aomori Prefecture are “Director-General of the Tourism and International Strategy Bureau.” The first words in Gunma Prefecture are “Director-General of the Transportation Infrastructure Department.” The first words in Shizuoka Prefecture are “Director-General of the Transportation Infrastructure Department.” From these examples, we can identify the primary political subjects in each prefecture at that time.

However, this methods has some drawbacks such as incorrect compounding. We can perform preliminary comparative analysis of political keywords using a simple tf-idf-weighting method. Further study will be required, *e.g.* statistical analysis such as corresponding analysis and machine learning methods.

4 Related work

Recently, some studies have explored document analysis, sentiment analysis, and political debates with respect to politics (Yano et al., 2009; Chambers et al., 2015; Cano-Basave et al., 2016). These studies

Table 4: Extracted political keywords without Compounding

Number	Prefecture	1	2	3
1	Hokkaido	HAC	Shigatani	Ainu culture
2	Aomori	Mutsu bay	NOVA	Towada city
3	Iwate	Uchimaru, Morioka city	Odashima	Daisetsu Ribaa-netto
4	Miyagi	Teizan canal	Sendai beef	privatization of Sendai airport
5	Akita	Kita-akita city	Ani Bear Farm	Area Nakaichi
6	Yamagata	Mogami-oguni River	Papua state	Tendo city
7	Fukushima	total inspection	South Aizu	return of the residents
8	Ibaraki	National Athletic Meet at Ibaraki	Tsuchiura city	Ushiku swamp
9	Tochigi	Tochimaru-kun	Haga district	Haga Red Cross Hospital
10	Gunma	Gunma's	Gunma Prefectural Assembly	Tomo wide-area arterial road
11	Saitama	Sasshin-no-kai	Saitama red cross hospital	members of Saitama Prefectural Assembly
12	Chiba	Lake Inba	Hokuso railway	Kujukuri
13	Tokyo	metropolitan government-owned land	an honorary citizen of Tokyo	Takaaki
14	Kanagawa	Sagami Longitudinal Expressway	KAST	Teruhiko
15	Niigata	Shibata	Shibata city	Niigata-Higashi Port
16	Toyama	members of Toyama Prefectural Assembly	kitokito	Imizu city
17	Ishikawa	Noto Satoyama Kaido	Noto Railway	Gyokusen'in
18	Fukui	Kuzuryu River	Asuwagawa Dam	Fukui Port
19	Yamanashi	Minami-Alps city	Kofu castle	Fujiyoshida city
20	Nagano	Asakawa Dam	Matsumoto Airport	NAGANO
21	Gifu	Mirai hall	Uchigatani	FC Gifu
22	Shizuoka	Granship	Yaizu fishing port	Numazu station
23	Aichi	Aichi Triennale	Linimo	Nishimikawa
24	Mie	Shinsei-Mie	AMIC	Odai town
25	Shiga	Omi Ohashi bridge	Mother Lake	Omihachiman city
26	Kyoto	Kyoto Stadium	Muko city	Nishijin Textile
27	Osaka	OTK	Neyagawa	Semboku Rapid Railway
28	Hyogo	Hometown, Hyogo	Muko River	Kobe Electric Railway
29	Nara	Nara City Hospital	Yamatotakada city	number 15, Mori-
30	Wakayama	Nisaka Yoshinobu	Obana Masahiro	Susami
31	Tottori	Lake Koyama-ike	Kurahoshi city	Wakasa town
32	Shimane	Sanko line	Shimane Japanese beef cattle	Kijima Dam
33	Okayama	Lake Kojima	Okayama marathon	Kurashiki station
34	Hiroshima	Yuzaki Hidehiko	Kaita town	members of Hiroshima Prefectural Assembly
35	Yamaguchi	Suo-oshima Island	Mitajiri-Nakanoseki Port	Shunan area
36	Tokushima	Mima city	Awa city	Miyoshi city
37	Kagawa	rare sugar	Kagawa Canal	All
38	Ehime	Mican	whole	Mikame town
39	Kochi	Kochi Junior College	Eikoku-ji Temple	Sukumo city
40	Fukuoka	Mount Hiko	Yabe River system	Okinohata River
41	Saga	Jobaru River	Ogi city	Saga Sakura Marathon
42	Nagasaki	Saikai city	Tsukumo	NERC
43	Kumamoto	Japanese brown cattle	Kumamoto urban area	Rogi Dam
44	Oita	Usa city	Hot Spring Prefecture Oita	Trinita
45	Miyazaki	Hososhima Port	Miyazaki Hospital	Toi cape
46	Kagoshima	Matsuza-	Dolphine Port	Marine Port Kagoshima
47	Okinawa	Shimojishima Airport	Nakagusuku Bay	Okinawan

used various documents as political corpora. In this section, we describe corpora that include political information.

Political document analysis has employed various document collections on the web, such as blogs (Yano et al., 2009). Probabilistic models have been proposed to generate blog posts and comments jointly within a blog site. Hassanali et al. (2010) proposed a technique to automatically tag political blog posts using support vector machines and named entity recognition. They used blog documents as a corpus. Chambers et al. (2015) modeled sentiment analysis for social sciences. They used Twitter data (over two billion tweets) as a corpus.Lerman et al. (2008) automatically predicted the impact of news on public perception of political candidates. They used daily newspaper articles as a corpus. Cano-Basave et al. (2016) used semantic frames to model argumentation in speaker discourse. Their presidential political debates corpus comprises 20 debates that took place between May 2011 and February 2012. Iyyer et al. (2014) applied a recursive neural network framework to detect political positions. They performed experiments using a dataset of Congressional debates and an original political dataset as a corpus. As mentioned above, political corpora typically comprise blogs, Twitter data, newspaper articles, and original political datasets. Therefore, a political corpus constructed from local assembly minutes is a novel and valuable source of political information.

5 Conclusion

In this paper, we described a Japanese political corpus constructed from the local assembly minutes of 47 prefectures collected over four years (April 2011 to March 2015), which represents a full term of office for assembly members in most autonomies. We collected local assembly minutes from all 47 prefectural assemblies, obtained fundamental statistical data from the collected text data, and extracted political keywords using a tf-idf weighting method.

As a result, we confirmed the following. First, we could collect 47 local assembly minutes for four years. Second, we could not completely divide the body text into "Name of speakers" and "Utterance" because the delimiter differs depending on the various autonomies. Finally, we demonstrated that the system can automatically extract political keyword using a tf-idf weighting method. We believe that this new corpus will be useful for comparative studies of local politics.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number JP16H02912.

References

Cano-Basave, Amparo Elizabeth and He, Yulan. *A Study of the Impact of Persuasive Argumentation in Political Debates*, Proceedings of NAACL-HLT, pp.1405–1413, 2016.

Chambers, N., Bowen, V., Genco, E., Tian, X., Young, E., Harihara, G., and Yang, E. *Identifying political sentiment between nation states with social media*, Proceedings of EMNLP, pp. 65–75, 2015.

Hassanali, Khairun-nisa, and Vasileios Hatzivassiloglou, *Automatic detection of tags for political blogs.*, Proceedings of the NAACL HLT 2010 Workshop on Computational Linguistics in a World of Social Media. Association for Computational Linguistics. pp. 21–22, 2010.

Iyyer, Mohit and Enns, Peter and Boyd-Graber, Jordan and Resnik, Philip. *Political ideology detection using recursive neural networks*, Proceedings of the Association for Computational Linguistics, 2014.

Kudo,Taku and Yamamoto, Kaoru and Matsumoto, Yuji, *Applying Conditional Random Fields to Japanese Morphological Analysis*, Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing (EMNLP-2004), pp.230-237, 2004.

Lerman, Kevin and Gilder, Ari and Dredze, Mark and Pereira, Fernando. Association for Computational Linguistics. *Reading the markets: Forecasting public opinion of political candidates by news analysis*, Proceedings of the 22nd International Conference on Computational Linguistics Vol. 1, pp.473–480, 2008.

Masuda, Tadashi , *Text Mining Analysis on the Minutes of Local Assemblies - A Case Study on the Takasaki City Assembly - (in Japanese)* . Takasaki City University Economics, Vol. 15, No.1, pp. 17–31, 2012.

Salton, Gerard and Buckley, Christopher, *Term-weighting approaches in automatic text retrieval*, Information processing & management Vol. 24, No 5, pp. 513–523, 1988.

Yano, Tae and Cohen, William W and Smith, Noah A. *Predicting response to political blog posts with topic models*, Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, pp.477–485, 2009.

Table 5: Extracted political keywords with Compounding

Number	Prefecture	1	2	3
1	Hokkaido	HAC	prefectural residents' life	Tomari power plant
2	Aomori	District Administration Office	Director-General of the Department of Planning and Polices	Director-General of the Tourism and International Strategy Bureau
3	Iwate	rules of Iwate Prefectural Assembly	big earthquake and tsunami	adding
4	Miyagi	ecological life, agriculture, forestry and fisheries	the -th Miyagi Prefectural Assembly	rules of Miyagi Prefectural Assembly
5	Akita	Chairperson of the Education and Public Safety Committee	Kitaakita city	hometown Akita cheering up plan
6	Yamagata	within the committee's jurisdiction	general branch office	Yamagata prefecture general accounting supplementary budget for - fiscal year
7	Fukushima	to hand out another paper	the Superintendent of Education	a report submitted by Chairperson
8	Ibaraki	rules of Ibaraki Prefectural Assembly	Mr. Onodera, the Superintendent of Education, takes the rostrum.	Ibaraki Liberal Democratic Party Branch
9	Tochigi	Chairperson of Tochigi Prefectural Assembly	rules of Tochigi Prefectural Assembly	Tochigi Genki Plan
10	Gunma	Director-General of the Life Culture and Sports Department	Director-General of the Citizens and Cultural Affairs Department	maximum time limit for the speech
11	Saitama	proposed bills	Director-General of the Department of Crisis Management and Disaster Prevention	Manager of the Sewerage Works
12	Chiba	Hokuso Railway	Eastern Chiba Medical Center	Kita-Chiba Road
13	Tokyo	following idea	examination of a petition	to meet an objective
14	Kanagawa	emergency financial measures	Director-General of the Public Health and Welfare Bureau	Chairperson of Kanagawa Prefectural Assembly
15	Niigata	the -th proposed bill	you all	Uonuma Kikan Hospital
16	Toyama	Chairperson of Toyama Prefectural Assembly	members of Toyama Prefectural Assembly	submit something with
17	Ishikawa	Noto Satoyama Kaido	Noto Railway	Director-General of the Citizens and Cultural Affairs Bureau
18	Fukui	Energy Research & Development Centralization Plan	examination report of a petition	Yosokichi
19	Yamanashi	Chairperson of Yamanashi Prefectural Assembly	rules of Yamanashi Prefectural Assembly	- minutes speech
20	Nagano	to declare the agenda	general administrative work	to declare the schedule
21	Gifu	country Gifu	supplementing of the budget	Director-General of the Urban Architecture Department
22	Shizuoka	Director-General of the Transportation Infrastructure Department	Director-General of the Economy, Trade and Industry Department	Shizuoka Prefectural Audit Commissioners
23	Aichi	Aichi and Nagoya	priority reform program	Aichi prefecture general accounting supplementary budget for - fiscal year
24	Mie	vision of citizens power	Shinsei-Mie	Higashikisyu
25	Shiga	Director-General of the Department of Public Works and Transportation	Director-General of the Department of Lake Biwa and the Environment	Chairperson of Shiga Prefectural Assembly
26	Kyoto	festivals and events	kissing loach	management of the commission in general
27	Osaka	Osaka Prefectural Government Sakishima Building	made in Osaka	OTK
28	Hyogo	Governor Ido takes the rostrum.	Kenmin Rengo Prefectural Assembly members	Hyogo Development
29	Nara	Prefectural Nara Hospital	bills for the year	Mount Wakakusa
30	Wakayama	Mr. Yoshinobu Nisaka	Mr. Yoshinobu Nisaka, Governor	Mr. Obana Masahiro
31	Tottori	Kurayoshi city	Manga Expo	Lake Koyama-ike
32	Shimane	Minshu-Kenmin Club	Mr. Mishima, a member of a Prefectural Assembly	the agenda
33	Okayama	Director-General of the Department of Environment and Culture	Mr. Kenro	Mr. Hideki
34	Hiroshima	Mr. Hidehiko Yuzaki, Governor	bills brought up together for discussion	explainer in charge
35	Yamaguchi	involved counselor	Maguchi Industrial Strategy Promotion Plan	Industry Strategy Headquarters
36	Tokushima	full name	be unique to Tokushima	A member attends the Prefectural Assembly.
37	Kagawa	rare sugar	Mr. Keizo Hamada, Governor	Kagawa Prefectural Assembly Minutes
38	Ehime	passage and verdict	Ehime National Sports Festival	Everyone raise their hands (in agreement).
39	Kochi	opinion and bill	Director-General of the Community Welfare and Services Department	Kochi Junior College
40	Fukuoka	within the -th bill's jurisdiction	bills brought up together for discussion	Chikuho area
41	Saga	Chairperson of Saga Prefectural Assembly	Director-General of the Transport Policy Department	Director of the Health and Welfare Headquarters
42	Nagasaki	Chairperson of the committee	Mr. Magome, a committee member	Mr. Yamaguchi, Chairperson
43	Kumamoto	Chairperson of Kumamoto Prefectural Assembly	rules of Kumamoto Prefectural Assembly	Tateno Dam
44	Oita	Mr. Nonaka, the Superintendent of Education	waiting seat	beside a facing podium
45	Miyazaki	bills proposed by a member of a Prefectural Assembly	Hososhima Port	Director-General of the Citizens Policy Department
46	Kagoshima	Kagoshima	Matsuza-	order of votes
47	Okinawa	We would like to take a break.	Shimojishima Airport	Council for Promotion of Dezoning and Reutilization of Military Land in Okinawa