東京都議会の会派を用いたStance classificationの試み

木村泰知[1]　佐々木稔[2]

[1] 小樽商科大学　[2] 茨城大学

kimura@res.otaru-uc.ac.jp, minoru.sasaki.01@vc.ibaraki.ac.jp

1 はじめに

近年，政治情報の課題に対して，自然言語処理を用いて解決するアプローチが期待されている．政治情報には，フェイクニュースの検出，フィルターバブルなどの課題がある．それらの課題に対しては，根拠のない意見や偏った意見をみつけることが必要となる．政治課題に関する意見には，政治のトピックを決めて，意見を判定する研究が行われている．

従来研究では，Twitter の投稿に対して意見を判定する研究が行われている．しかしながら，意見推定は，一つのトピックに対しての意見を判定しており，異なる複数のトピックを考慮していないことから，発言者の考え方を捉えることが難しいという問題がある．また，発言者の意図を捉えるためには，発言者の所属している組織が影響している．例えば，政治家の所属している政党，記者の所属している新聞社などの立場が発言に影響する．上記の意見抽出に関する問題に対しては，発言者の意見を考慮しつつ，発言者の立場を推定するアプローチがあり，Stance classification，あるいは，Stance detection と呼ばれる．発言者の立場は，一つのトピックに対する意見推定だけで判断できないことから，複数の政治トピックに対する意見から総合的に判断することになる．例えば，新聞社，政党，会派などの違いは，発言者の意見に影響を与えており，立場と呼ぶことができる．これらの研究は，複数トピックに対する意見を統合して捉えようとするアプローチといえる．

そこで，本研究では，東京都議会の会派を Stance として捉えた Stance classification を行う．本研究の目的は，東京都議会の会派を用いた Stance classification のタスクの提案であり，さらに，既存手法による性能評価を行うことで，タスクの難しさを明らかにすることである．図 1 に本研究におけるイメージ図を示す．本研究では，会派を Stance として分類することに加えて，選挙の前後のような時間軸により変化する会派の変化も含めて，Stance classification を行うことを試みる．

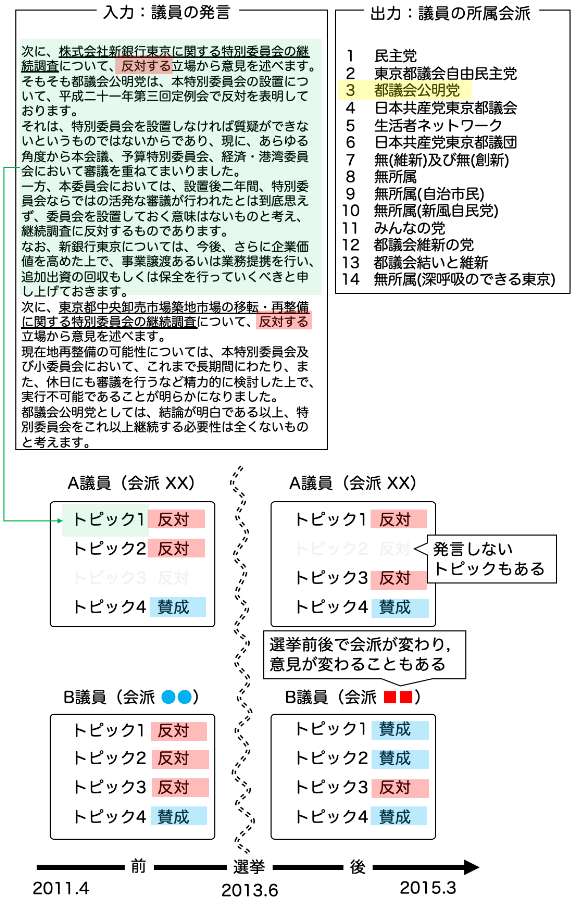

図 1: 本研究の Stance classification のイメージ

2 関連研究

Sentiment analysis と Stance Classification は，意見を対象とした分類問題という点で共通しているが，対象文中に明確な意見が必ず存在するのか，あるいは，必ずしも存在しないのかに違いがある [1].

Copyright(C) 2019 The Association for Natural Language Processing.
All Rights Reserved.

Sentiment analysis は，トピック単位についての Positive, Negative, あるいは Neutral を推定している．一方で，Stance Classification は，発言者，あるいは，投稿者の Stance(立場) を表しており，その発言者の信念といえる．SemEval-2016 Task-6 [3] は，Tweet から Stance を見つけるタスクである．例えば，入力に "Hillary Clinton" と "Legalization of Abortion" と Tweet の「The pregnant are more than walking incubators, and have rights!」が与えれら，スタンスとして favor，against，neither を推定する．また，Walker らは，ウェブサイトにあるフォーラムを対象とした Stance Classification をしている [5]． さらに，Stance classification に政党を利用する研究がある．しかしながら，政党は，分類するための素性としてのみ利用されている [2]. 従来の Stance classification では，2 値，あるいは，3 値の分類であり [4][5]，発言者の立場を 3 つのカテゴリに分類しづらい場面も存在する．

上記の問題点を踏まえて，本研究では，東京都議会における会派を Stance(立場) として Stance classification を行う．会派は，議会において代表質問をするために，当選した議員によって作られる．東京都議会の場合には，5 人以上の議員により構成された会派が代表質問をすることができる．そのため，会派は，政党と同じ名前の場合もあるが，独自の会派も数多く存在しており，地方議会議員の Stance をうまく表しているといえる．

3 データセット

3.1 作成方法

本研究では，東京都議会を対象として，議員の会派を Stance(立場) として，議員名，議員の発言，会派の 3 つ組のデータセットを作成する．東京都議会を対象とした理由として，東京都議会の議員の会派を収集しやすいことが挙げられる．東京都議会会議のウェブサイトには，議員の所属会派に加えて，開始と終了の期間が記載されている．

次に，対象期間について説明する．我々は，従来から，統一地方選挙に当選した議員が任期を終えるまでの 4 年間を対象として集めており，東京都議会会議録の対象期間についても同様である．しかしながら，東京都は，1965 年 6 月に東京都議会が解散されたことに伴って，統一地方選挙の時期とずれ，独自の選挙が行われるようになっている．このように，東京都は，統一地方選挙と異なるため，我々が収集したデータに含まれる 2013 年 6 月の「選挙前 (2011 年 4 月から 2013 年 5 月まで)」および「選挙後 (2013 年 6 月から 2015 年 3 月まで)」の時期をデータセットとすることにした．表 2 と表 3 に選挙前と選挙後のデータセットの議員数，単語数，平均単語数，単語数の中央値，最大単語数，最小単語数を示す．

次に，データセットの具体的な作成について述べる．本タスクは，入力が議員一人の発言であり，出力が議員の所属会派となる．学習データおよびテストデータの作成に当たり，両方のデータに 1 人以上の所属会派の議員が存在するようにデータセットを作成することとした．つまり，同一会派で 2 人以上発言している会派を対象として，半分を学習データ，もう半分をテストデータとして用いることとした．また，任期期間中に所属を変更した議員については，対象としないこととした．

3.2 東京都議会の会派

表 1 には，2011 年 4 月から 2015 年 3 月までに東京都議会で発言した議員を対象として「会派名」と「所属議員数」を記述している．2013 年 6 月に東京都議会選挙が行われたため「選挙前 (2011 年 4 月から 2013 年 5 月まで)」および「選挙後 (2013 年 6 月から 2015 年 3 月まで)」に分けて記述している．

表 1: 2013 年 6 月の「選挙前」と「選挙後」の会派名と発言した議員数

会派名	選挙前の議員数	選挙後の議員数
民主党	42	14
東京都議会自由民主党	30	51
都議会公明党	23	20
日本共産党東京都議会	6	12
生活者ネットワーク	2	2
日本共産党東京都議団	1	1
無 (維新) 及び無 (創新)	1	0
無所属	1	0
無所属 (自治市民)	1	0
無所属 (新風自民党)	1	1
みんなの党	0	3
都議会維新の党	0	1
都議会結いと維新	0	1
無所属 (深呼吸のできる東京)	0	1

4 実験

4.1 目的

本実験の目的は，東京都議会の会派を用いた Stance classification のタスクの提案であり，さらに，既存手

Copyright(C) 2019 The Association for Natural Language Processing. All Rights Reserved.

表 2: 選挙前：学習データとテストデータ の議員数，単語数，平均単語数，中央値，最大単語数，最小単語数.

会派	学習データ						テストデータ					
	議員数	単語数	平均値	中央値	最大値	最小値	議員数	単語数	平均値	中央値	最大値	最小値
民主党	21	61,048	2907.1	1,569	15,084	161	21	54,654	2,602.57	2,098	10,066	117
都議会公明党	12	22,926	1,910.5	1,858	3442	89	11	33889	3080.82	2,417	8,413	87
東京都議会自由民主党	15	32,054	2,136.9	1,614	4,870	508	15	54374	3,624.93	2,452	12,127	988
生活者ネットワーク	1	3,254	3,254	3,254	3,254	3,254	1	5,425	5,425	5,425	5,425	5,425
日本共産党東京都議会	3	11,178	3,726	3,272	4,870	3,036	3	12,455	4,151.67	5,015	5,308	2,132
無所属 (自治市民)	1	2,634	2,634	2,634	2,634	2,634	0	0	0	0	0	0
日本共産党東京都議団	1	3,739	3,739	3,739	3,739	3,739	0	0	0	0	0	0
無所属	1	8,975	8,975	8,975	8,975	8,975	0	0	0	0	0	0
無 (維新) 及び無 (創新)	1	1,120	1,120	1,120	1,120	1,120	0	0	0	0	0	0
無所属 (新風自民党)	1	1,009	1,009	1,009	1,009	1,009	0	0	0	0	0	0
合計	57						51					

表 3: 選挙後：学習データとテストデータ の議員数，単語数，平均単語数，中央値，最大単語数，最小単語数.

会派	学習データ						テストデータ					
	議員数	単語数	平均値	中央値	最大値	最小値	議員数	単語数	平均値	中央値	最大値	最小値
民主党	7	9,726	1,389.4	1,317	1,917	939	7	25,522	3,646	2,320	9,408	926
みんなの党	2	5,388	2,694	2,694	2,838	2,550	1	2,074	2,074	2,074	2,074	2,074
日本共産党東京都議会	6	25,208	4,201.3	4,118.5	5,463	3,312	6	12,071	2,011.83	1,876.5	2,873	1,584
東京都議会自由民主党	26	65,083	2,503.2	1,875	11,062	790	25	58,400	2,336	1,776	12,231	598
都議会公明党	10	2,2041	2,204.1	1,750.5	4,153	965	10	2,9437	2,943.7	2,225.5	8,261	570
生活者ネットワーク	1	4,020	4,020	4,020	4,020	4,020	1	3,256	3,256	3,256	3,256	3,256
都議会維新の党	1	1,606	1,606	1,606	1,606	1,606	0	0	0	0	0	0
無所属 (新風自民党)	1	2,912	2,912	2,912	2,912	2,912	0	0	0	0	0	0
無所属 (深呼吸のできる東京)	1	2,383	2,383	2,383	2,383	2,383	0	0	0	0	0	0
日本共産党東京都議団	1	3,413	3,413	3,413	3,413	3,413	0	0	0	0	0	0
都議会結いと維新	1	2,031	2,031	2,031	2,031	2,031	0	0	0	0	0	0
	57						50					

法による比較実験を行うことで，タスクの難しさを明らかにすることである．

4.2 方法

本実験では，東京都議会の選挙が行われた前と後の会議録を利用する．対象期間は，2011 年 4 月から 2015 年 3 月までであり，選挙前後に分けてデータを作成する．データセットは，3.1 で述べた通り，選挙前，および，選挙後の東京都議会会議録を議員ごとの発言に分けて作成する．各議員の発言は，一つにまとめており，議員名，会派の情報が付与されている．入力形式は，議員の発言に含まれる単語の頻度を要素とするベクトルである．単語の分割は，形態素解析器の MeCab (辞書に neologd) を用いている．利用する単語は，名詞，動詞，形容詞とした．比較手法は，SVM，決定木，k-NN，Random Forest，Naive Bayes である．

4.3 結果

実験結果は，2013 年 6 月の選挙の「前」と「後」に分けて，表 4 に示す．表 4 からわかるように，Random Forest が最も高い正解率であった．

表 4: 選挙前の実験結果

手法	Parameter	選挙前 正解率	選挙後 正解率
SVM	kernel=rbf	0.4902	0.52
決定木	深さ=3	0.5882	0.48
k-NN	k=3	0.4118	0.26
Random Forest	深さ=5 決定木の数=10	0.6078	0.56
Naive Bayes		0.4510	0.56

4.4 考察

SVM，Random Forest の正解と予測結果の confusion matrix から，手法による違いを明らかにする．選挙前の図 2 と図 3 や選挙後の図 4 と図 5 を比較するとわかるように，SVM は予測ラベルが一つの会派に集中しているが，Random Forest は一つの会派に集中していない．Random Forest は弱学習器を用いて判断する長所をうまく利用していると考えられる．

本実験では，任期期間内に会派を変えた議員を対象外とした．しかしながら，政党および会派を変えた場合にも追随することが Stance classification にも必要である．例えば「くりした善行」議員は「民主党」「東京維新の会」「都民ファーストの会」へと会派を変更している．このように政党および会派の方針に合わな

Copyright(C) 2019 The Association for Natural Language Processing.
All Rights Reserved.

い場合には，Stance を変更することがあり，議員の方針の変化を発言から判断できるようになることが必要となる．

	選挙前 SVM	予測したラベル						
			0	1	2	3	4	
正	民主党	0	21	0	0	0	0	21
解	都議会公明党	1	3	0	0	0	0	3
ラ	東京都議会自由民主党	2	13	0	2	0	0	15
ベ	生活者ネットワーク	3	8	0	1	2	0	11
ル	日本共産党東京都議会	4	1	0	0	0	0	1
			46	0	3	2	0	51

図 2: 選挙前の SVM

	選挙前 RF	予測したラベル							
			0	1	2	3	4	5	
正	民主党	0	16	0	3	2	0	0	21
解	都議会公明党	1	0	2	0	1	0	0	3
ラ	東京都議会自由民主党	2	6	0	7	1	0	1	15
ベ	生活者ネットワーク	3	2	0	3	6	0	0	11
ル	日本共産党東京都議会	4	0	1	0	0	0	0	1
	無所属(自治市民)	5	0	0	0	0	0	0	0
			24	3	13	10	0	1	51

図 3: 選挙前の Random forest

	選挙後SVM	予測したラベル							
			0	1	2	3	4	6	
正	民主党	0	0	0	0	7	0	0	7
解	みんなの党	1	0	0	0	1	0	0	1
ラ	日本共産党東京都議会	2	0	0	0	6	0	0	6
ベ	東京都議会自由民主党	3	0	0	0	25	0	0	25
ル	都議会公明党	4	1	0	0	8	1	0	10
	生活者ネットワーク	6	0	0	0	1	0	0	1
			1	0	0	48	1	0	50

図 4: 選挙後の SVM

	選挙後 RF	予測したラベル							
			0	1	2	3	4	6	
正	民主党	0	1	0	0	5	1	0	7
解	みんなの党	1	0	0	0	1	0	0	1
ラ	日本共産党東京都議会	2	0	0	1	4	1	0	6
ベ	東京都議会自由民主党	3	1	0	0	24	0	0	25
ル	都議会公明党	4	0	0	0	8	2	0	10
	生活者ネットワーク	6	0	0	0	1	0	0	1
			2	0	1	43	4	0	50

図 5: 選挙後の Random forest

5 おわりに

本稿では，東京都議会の会派を用いた Stance classification のタスクの提案を行った．また，既存手法による比較実験を行うことで，タスクの難しさを明らかにした．その結果，Random Forest が最も正解率が高い結果となり，現時点で約 6 割の正解率となるタスクであることを確認した．

今後は，議員の会派離脱にも考慮したデータセットの整備を進めるとともに，分類精度の向上を目指した手法の開発を行う．

謝辞

本研究は JSPS 科研費 JP16H02912 の助成を受けたものです．

参考文献

[1] Javid Ebrahimi, Dejing Dou, and Daniel Lowd. Weakly supervised tweet stance classification by relational bootstrapping. In *EMNLP 2016*, pages 1012–1017, 2016.

[2] Kenneth Joseph, Lisa Friedland, William Hobbs, David Lazer, and Oren Tsur. Constance: Modeling annotation contexts to improve stance classification. In *EMNLP 2017*, pages 1115–1124, 2017.

[3] Saif Mohammad, Svetlana Kiritchenko, Parinaz Sobhani, Xiaodan Zhu, and Colin Cherry. Semeval-2016 task 6: Detecting stance in tweets. In *SemEval-2016*, pages 31–41, June 2016.

[4] Dhanya Sridhar, Lise Getoor, and Marilyn Walker. Collective stance classification of posts in online debate forums. In *Workshop of ACL 2014*, pages 109–117, June 2014.

[5] Marilyn A. Walker, Pranav Anand, Robert Abbott, and Ricky Grant. Stance classification using dialogic properties of persuasion. In *NAACL HLT '12*, NAACL HLT '12, pages 592–596, 2012.

Copyright(C) 2019 The Association for Natural Language Processing.
All Rights Reserved.