



筑波大学
University of Tsukuba

内閣支持率と答弁データの 動的相関の研究

樋口心

視覚メディア研究室

目次

- スライドの内容
 - ◆ テーマ全体の説明
 - ◆ 今回の進捗説明
 - ◆ 結果
 - ◆ 考察

テーマ

「内閣支持率と答弁データの動的相関の研究」

研究目的

内閣支持率と答弁言語データの時系列解析

従来研究では、データを測定した時点での特徴しか得られない。

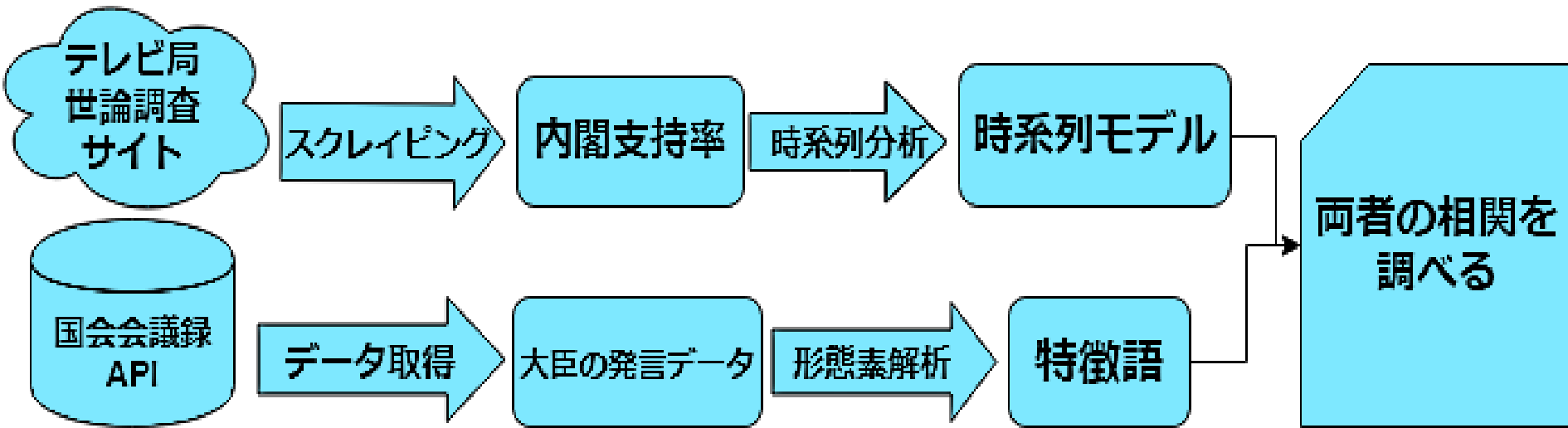


時系列データと比較することにより、
動的に言語データをグルーピングする。



言語データの動的分析手法の確立

研究手法(全体)

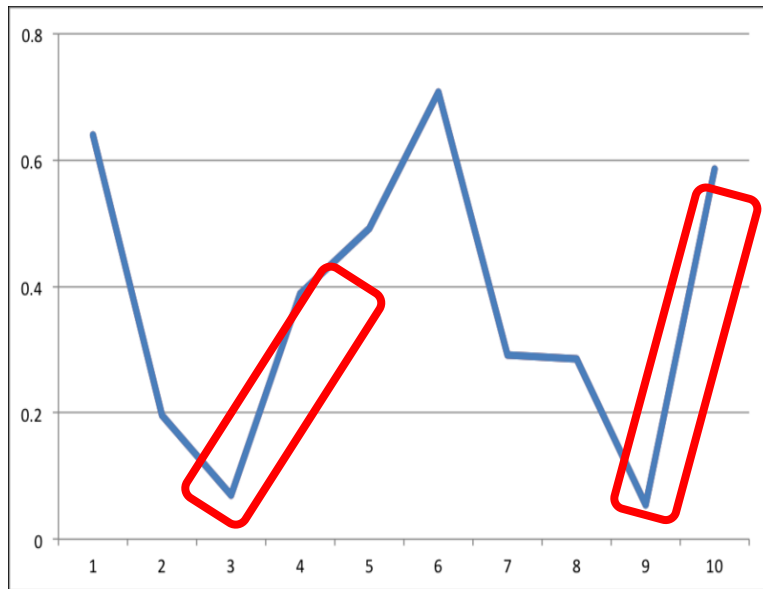


1. 形態素解析を行い、大臣の発言データを品詞ごとに分類
2. 内閣支持率データを時系列解析し、変動が大きい部分に着目

両データを比較し、有益な相関を見出す。

研究手法②(詳細)

内閣支持率データ



支持率が上がる**直前**に出てくる
特徴語

支持率が上がっている**最中**に
出てくる特徴語

支持率が上がった**後**に出てくる
特徴語

etc.

支持率の変動によって言語データをグループ分けし、
支持率の上下により、発言に特徴が出るか検証する

今回の進捗内容

- データの収集
 - 月毎に安倍総理大臣の答弁データを集める
 - NHKのサイトから内閣支持率を収集
- データのグルーピング
 - 支持率が上下するタイミング毎にデータを纏める
- 両データの特徴分析
 - 接続詞の順接・逆接の使用頻度の確認
 - 頻出単語を確認し、特徴語を分析
 - 機械学習を行い、支持率の上下により言語データを判別できるか

収集データの詳細

	国会会議録API	内閣支持率
取得時期	2013/01~2018/8	2013/01~2018/8
取得対象者/取得元	安倍晋三	NHK
使用素性	動詞・形容詞・接続詞	-
データセット	月毎に一つにまとめる	-

発言データが存在しない月

↓	2015年12月
2013年7月	2016年6月
2013年8月	2016年7月
2013年9月	2016年8月
2014年8月	2017年8月
2014年12月	2017年9月
2015年10月	2017年10月

データのグルーピング方法

	2013/1/1	2013/2/1	2013/3/1	2013/4/1	2013/5/1	2013/6/1
支持する	64	64	66	66	65	62
支持率前月比（遅行係数）	-	0	2	0	-1	-3
支持率当月比（現在係数）	0	2	0	-1	-3	-5
支持率翌月比（先行係数）	2	0	-1	-3	-5	0

支持率の上下により各月の係数を決定

遅行係数…支持率が上下した値を翌月の発言につける

一致係数…上下した値を当月の発言につける

先行係数…上下した値を前月の発言につける



各係数が上昇・降下した月のデータを値が大きい順に10個をデータセットとした

データのグルーピング方法②

現在係数	
2017/10/1	9
2014/8/1	7
2018/6/1	6
2015/8/1	6
2017/8/1	5
2016/12/1	5
2016/10/1	5
2016/7/1	5
2015/3/1	5
2017/7/1	4
2016/8/1	4
2016/1/1	4
2015/10/1	4
2015/1/1	4
2014/7/1	4
2014/4/1	4
2013/12/1	4
2017/11/1	3
2017/1/1	3
2016/5/1	3
2016/4/1	3

2017/6/1	-13
2013/11/1	-10
2014/10/1	-8
2015/2/1	-8
2015/6/1	-7
2016/9/1	-7
2017/2/1	-7
2017/9/1	-7
2014/9/1	-6
2018/3/1	-6
2013/6/1	-5

1. 各係数、昇順/降順に並べる
2. 存在しないデータ以外で10か月を集め
1つのデータセットとする
3. 6つのデータセットに分類する

- 現在減
- 現在増
- 先行減
- 先行増
- 遅行減
- 遅行増

目次

●今回の内容

- ◆テーマ全体の説明

- ◆今回の進捗説明

◆結果

1. 逆接/順接の頻度検証
2. 各データセットの特徴語
3. 機械学習による分類

- ◆考察

1. 逆接/順接の頻度検証

● 仮定

支持率が下がる時には野党の質問を
繰り返すために、
「しかし」「ですが」などの逆接の表現が増
えるのではないか

● 結果

先行係数・遅行係数では、
逆接の表現が増えていることが判明

現在減	現在増	先行減	先行増	遅行減	遅行増
0.805	0.828	0.867	0.841	0.810	0.778

しかし、決定的な値とは言えない・・・？

順接	現在減	現在増	先行減	先行増	遅行減	遅行増
だから	34	40	46	31	63	18
それで	14	8	19	16	20	12
そこで	424	390	441	403	544	277
したがって	44	41	37	55	57	35
ゆえに	4	3	2	5	6	8
すると	0	1	0	0	0	0
それなら	1	0	0	0	0	0
それでは	7	9	11	8	9	6
それで	14	8	19	16	20	12
合計	542	500	575	534	719	368

逆接	現在減	現在増	先行減	先行増	遅行減	遅行増
しかし	594	536	597	568	808	419
しかしながら	22	14	15	23	14	25
けれども	0	0	0	1	1	0
ところが	4	2	3	2	2	2
それなのに	0	0	1	0	1	0
でも	43	33	38	28	47	20
それでも	10	19	9	13	15	7
合計	673	604	663	635	888	473
比率	0.805	0.828	0.867	0.841	0.810	0.778

2. 各データセットの特徴語(動詞)

支持率が上がった時に多くみられる動詞

先行		現在		遅行	
知る	3.7	与える	1.5	まいる	2.4
増える	2.4	呼ぶ	1.5	応える	2.3
満たす	2.4	認める	1.4	引き続く	2.1
応える	2.1	取る	1.4	目指す	1.7
至る	1.9	脅かす	1.4	向ける	1.6
取る	1.7	至る	1.4	取り組む	1.6
脅かす	1.6	おく	1.3	努める	1.6
果たす	1.6	入る	1.3	続ける	1.4
繰り返す	1.5	見る	1.3	進める	1.4
尽くす	1.5	守る	1.3	させる	1.4

使用頻度上位100
単語の中から選定

支持率が下がった時に多くみられる動詞

とる	1.7	応える	1.8	書く	2.2
呼ぶ	1.6	起こる	1.8	知る	2.1
させる	1.6	わかる	1.8	関わる	2.1
おっしゃる	1.5	しれる	1.6	満たす	2.0
かかわる	1.4	もらう	1.6	守る	2.0
上がる	1.4	関わる	1.6	起こる	1.8
いける	1.4	かかわる	1.5	来る	1.6
行く	1.4	生かす	1.5	行く	1.6
やる	1.4	変える	1.4	あり	1.6
踏まえる	1.4	尽くす	1.4	重ねる	1.5

2. 各データセットの特徴語(形容詞)

支持率が上がった時に多く見られる形容詞

先行	
危うい	8.5
おいしい	3.3
固い	2.5
明るい	2.5
ほしい	2.4
粘り強い	2.1
にくい	1.8
激しい	1.8
弱い	1.5
乏しい	1.4

現在	
安い	3.8
やむを得ない	2.8
著しい	2.1
おかしい	1.9
長い	1.8
弱い	1.7
深い	1.7
良い	1.7
正しい	1.6
新しい	1.5

遅行	
粘り強い	6.6
おいしい	5.5
乏しい	4.4
きめ細かい	3.3
著しい	2.6
固い	2.6
幅広い	2.2
がたい	2.2
明るい	2.0
温かい	2.0

使用頻度上位50
単語の中から選定

支持率が下がった時に多く見られる形容詞

おかしい	2.9
長い	2.2
短い	2.0
深い	1.9
著しい	1.6
近い	1.5
難しい	1.5
少ない	1.4
いい	1.4
低い	1.3

すばらしい	3.8
乏しい	3.3
危うい	2.5
温かい	1.9
明るい	1.7
激しい	1.7
短い	1.6
がたい	1.5
近い	1.5
力強い	1.5

ほしい	3.8
詳しい	2.1
近い	2.1
やむを得ない	2.1
危うい	2.0
にくい	1.8
いい	1.7
短い	1.6
良い	1.6
おかしい	1.6

特徴語がつかわれている文脈

片方に寄っている特徴語をキーワードを実際に国会会議録内で検索しました。

短い

- …「短いメッセージ」→籠池氏からの留守電メッセージ
- …「短い期間でデフレ脱却」→デフレ脱却・アベノミクス

粘り強い

- …「8年間に及ぶ粘り強い交渉」→TPP

おいしい

- …「おいしいお米/魚」→食文化・TPP・日本の農水産物の世界展開

3. 機械学習による分類

2016-03-01.txt	474 KB
2016-04-01.txt	108 KB
2016-05-01.txt	75 KB



2016-03-01_1.txt	72 KB
2016-03-01_2.txt	83 KB
2016-03-01_3.txt	88 KB
2016-03-01_4.txt	115 KB
2016-03-01_5.txt	52 KB
2016-03-01_6.txt	66 KB
2016-04-01_1.txt	108 KB
2016-05-01_1.txt	75 KB

100発言毎にファイルを別にすることで、

- ・ データセット毎のデータ量の偏り
- ・ データセットが少ないこと

を解消

データセット	支持率上昇	支持率降下
ラベル数	24	24
データ量	2.15 MB	2.53 MB
使用素性	動詞/形容詞	動詞/形容詞

結果

動詞

	precision	recall	f1-score	support
支持率上昇	0.61	0.46	0.52	24
支持率降下	0.57	0.71	0.63	24
avg / total	0.59	0.58	0.58	48

	0	1
0	11	13
1	7	17
	predicted label	

形容詞

	precision	recall	f1-score	support
支持率上昇	0.60	0.50	0.55	24
支持率降下	0.57	0.67	0.62	24
avg / total	0.59	0.58	0.58	48

	0	1
0	12	12
1	8	16
	predicted label	

F値、58%と低い精度

考察・今後の方針

- 同じ人の発言のため、特徴が出にくい。

→したがって、機械学習における判別が難しい

- Bag of Wordsの次元が大きすぎるため過学習してしまう。

→LSI等による次元削減を試みる。

→TF・IDF値を用いてみる

→出現頻度が高すぎる/低すぎる単語をあらかじめ取り除く。

- トピック生成など、特徴がつかみやすい形に出力する

→ただ、頻出単語などで見ても特徴がつかみづらいので、グループ分けに工夫する&特徴が出やすい元データの収集をしないと難しいと思われる。