

Politicians in Youtube: How do they use their Youtube Channel?

Hyun Woo Lim

WUSTL Political science

- Youtube is one of the major information sources
- Little known about how legislators show themselves in their Youtube Channel
- Video has many features including text, image and audio

Scholars have studied political advertising in Youtube, but they have not focused how legislators use their Youtube channels overall.

Systematic analysis of large scale Youtube data is not done.

- Klotz 2010; Lev-On 2012; Borah, Porisma, Erika Fowler, and Ridout 2018

Goals: data

Short term goal

- Find descriptive statistics: channel data and video data
 - video.id, post.time, title, duration, view.count, like.count, dislike.count, comment, transcript etc

Long term goal

- download video, download transcript for analysis

Goals: research questions

Short term goal

- Which politicians use Youtube for what purposes?
- Are politicians' Youtube contents different from their other media outlets such as Twitter and Facebook?

Long term goal

- How do politicians use emotions to affect audience?
- How do politicians use images?
- What is politicians' campaign strategy?

Data collecting process

Video and transcript

- Using all legislators' available Youtube channel id, I scrapped the videos urls.
- Download data to local drive and convert them
- Move the data to Box cloud drive
- Using Cluster to finish the looping

Descriptive data of videos

- Youtube API
- Other Python packages

Goal: All legislators' available Youtube data from their channels

Current status:

- Total 484 legislators (House and Senate) as of 2018 October
- 14265 URLs
- 5871 videos
- 5155 captions
- Total about 300 GB

Outline of processed data

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	title_mat	caption	url	title_ori	message	duration	author	length	thumb	videoid	viewcount	likes	dislikes	
2	011317Col		https://w	01/13/17	Cspan	0:03:40	Doug Colli	220	http://i.yt	sIDB6r9Dn	96	2	0	
3	01172017C		https://w	01/17/2017	Congress	0:03:47	Doug Colli	227	http://i.yt	iRsl-tf84n	85	0	0	
4	011818Col		https://w	01 18 18	Collins Joins	0:05:41	Doug Colli	341	http://i.yt	aAnX8Ggt	69	0	0	
5	021318Col		https://w	02 13 18	Collins Talks	0:04:40	Doug Colli	280	http://i.yt	QjkTJU4	56	3	1	
6	022314Rep		https://w	02.23.14:	Rep. David	0:02:32	Lauren Bu	152	http://i.yt	JIN2rlqUth	116	1	0	
7	030911Sta		https://w	03.09.11	State and Fo	0:05:29	RepKayGr	329	http://i.yt	hLH5Qjh-t	1016	1	0	
8	033011Sta		https://w	03.30.11	State and Fo	0:10:57	RepKayGr	657	http://i.yt	KKsEMeJC	135	1	0	
9	051816Col		https://w	05/18/16	Collins Def	0:01:37	Doug Colli	97	http://i.yt	Rx2xiyClfI	163	1	0	
10	051915Cla		https://w	05 19 15	Clay VIDEO	0:00:26	Congressr	26	http://i.yt	Z7_zoK7W	53	0	0	
11	062411Rep		https://w	06.24.11:	Rep David S	0:01:25	Lauren Bu	85	http://i.yt	PBNDZ0nJ	58	0	2	
12	062718Col		https://w	06 27 18	Collins Discu	0:03:16	Doug Colli	196	http://i.yt	sYarLkQ4p	104	2	0	
13	062718Col		https://w	06 27 18	Collins Discu	0:02:13	Doug Colli	133	http://i.yt	WW9B9Pc	157	3	0	
14	06272018F		https://w	06 27 2018	Rep. Vela	0:05:10	Nydia Vel	310	http://i.yt	jzFQd12e2	51	3	2	
15	071118Col		https://w	07 11 18	Collins Discu	0:04:40	Doug Colli	280	http://i.yt	CvC-jaZog	546	13	0	
16	072318Col		https://w	07 23 18	Ci newsinter	0:04:27	Doug Colli	267	http://i.yt	IULRqPXIC	133	17	0	
17	072318Col		https://w	07 23 18	Collins Resp	0:04:37	Doug Colli	277	http://i.yt	nvx_tjK6Y	70	15	0	
18	072418Col		https://w	07 24 18	Collins on N	0:04:31	Doug Colli	271	http://i.yt	_wzbM4iU	54	8	0	
19	07716Colli		https://w	07/7/16	Collins: Abu	0:05:52	Doug Colli	352	http://i.yt	rjmFILjOEI	238	6	0	
20	090418Col		https://w	09 04 18	Collins Com	0:04:03	Doug Colli	243	http://i.yt	AWyLxHe'	174	11	0	
21	090518Enc		https://w	09 05 18	Energy and	0:04:28	RepGusBil	268	http://i.yt	thMlUwA	79	0	0	

Data: transcript (caption)

Rep. House Todd Young (current Senator) on Gun control [▶ Link](#)

Transcript:

hello and welcome to our latest edition of talking with Todd an opportunity for me to periodically share with you my thoughts on important issues before Congress and to discuss topics which are important to Hoosiers today I'll focus on some of the proposals to reduce gun violence that have been discussed in recent weeks many of you have called or in the office to ask where I stand thank you let's not mince words I'm a strong supporter of the Second Amendment and I have an A rating with the National Rifle Association while I can support measures to keep guns out of the hands of criminals I will not support any legislation that infringes on the ability of law-abiding citizens to exercise their constitutionally guaranteed right to the legal gun ownership the Supreme Court has made it clear that the Second Amendment right to keep and bear arms is an explicitly

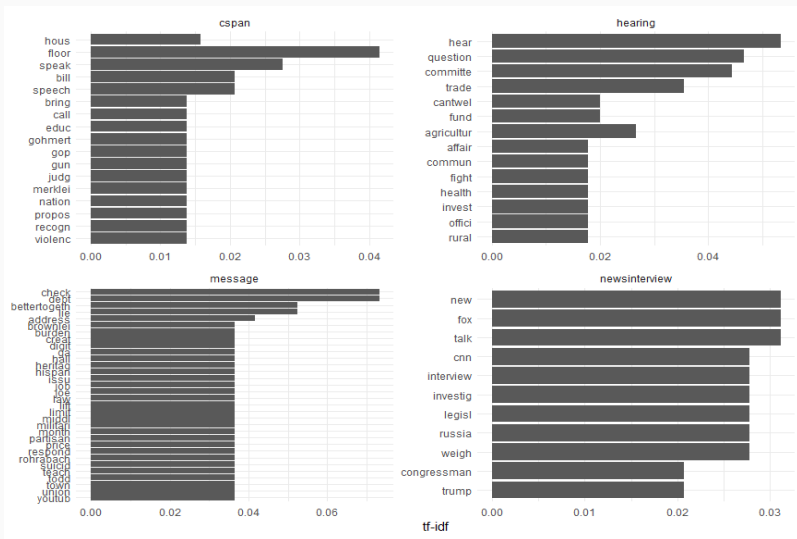
Method for classification

Goal: Classifying types of videos

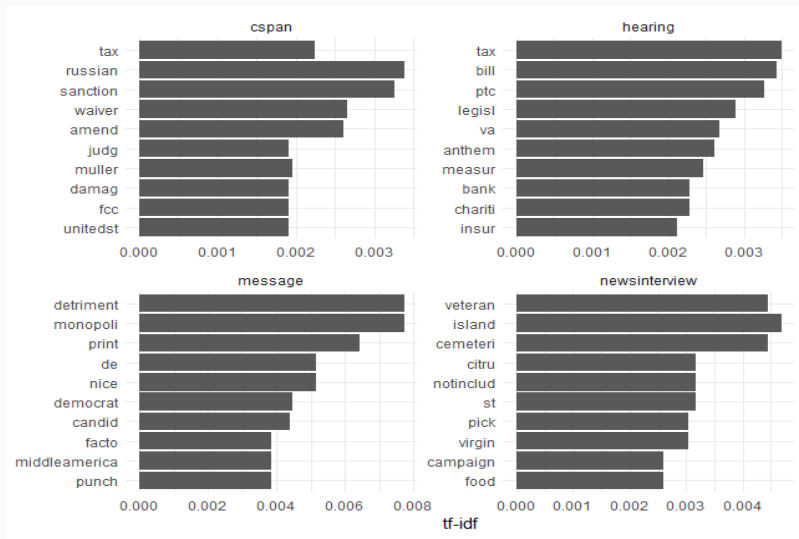
Method: supervised learning: random forest

- Hand-coded 200 videos: 9 types
 - cspan, hearing, message, mis, news, newsinterview, press, service, speech
- Create document-term matrix
- Exploratory analysis: chart
- Train the model (using video title or transcript)
- Test the model

Exploratory analysis: using title



Exploratory analysis: using transcript



Prediction: using title

35 percent accuracy

	preds									
	cspan	hearing	message	mis	news	newsinterview	press	service	speech	
cspan	12	10	2	2	0	9	1	4	1	
hearing	2	9	1	1	0	3	0	3	0	
message	2	0	3	0	1	3	0	3	0	
mis	1	0	2	0	1	0	0	0	0	
news	0	0	0	0	0	3	0	0	0	
newsinterview	1	1	0	1	1	10	0	0	1	
press	0	0	0	0	0	0	1	0	0	
service	1	0	0	0	0	0	0	0	0	
speech	1	3	0	0	0	0	0	0	0	

Prediction: using transcript

31 percent accuracy

	preds						
	cspan	hearing	message	mis	news	newsinterview	speech
cspan	30	0	0	0	0	1	0
hearing	23	1	0	0	0	1	0
message	15	1	0	0	0	1	0
mis	2	0	0	0	0	0	0
news	1	0	0	0	0	0	0
newsinterview	14	1	0	0	0	0	0
press	4	0	0	0	0	0	0
service	2	0	0	0	0	0	0
speech	3	0	0	0	0	0	0

Conclusion

- Using title shows better accuracy than using transcript based on random forest model.
- Random forest model may not be the best model.
- There are problems in hand-coding.
 - Too many different classifier and subtle differences.
 - Only 100 train data set and 100 test data set
- Unsupervised learning can work better.

Future direction

Data

- Scrap all videos from given Channels
- Use unsupervised learning models
- Get comments data from each videos

Research questions

- Examine which legislators use Youtube channel for what purposes
- Study how electoral contexts affect their Youtube videos
- Study strategic usage of emotion, image and text from videos