

# Politicians in Youtube: How do they use their Youtube Channel?\*

Hyun Woo Lim †

Washington University in St. Louis

April 4, 2020

---

\*Notes: Author thanks Chris Lucas for advice and help.

†Ph.D. Student, Department of Political Science, Campus Box 1063, One Brookings Drive, St. Louis, MO 63130; [hyunwoolim@wustl.edu](mailto:hyunwoolim@wustl.edu).

# **Introduction**

How do politicians use their Youtube channels? Political scientists have studied how legislators use their Twitter and Facebook, but their Youtube channels have not been investigated by political scientists. Although Youtube is one of the major information source for today's citizens, politicians' Youtube channels have not been explored probably for mainly two reasons: First, Youtube data is not easily accessible compared to other media outlet. Even collecting the complete list of Youtube videos from a channel was not easy to do. Second, analyzing video data is not easy. Video is resourceful in that it has image, audio and text components, but analyzing them is difficult.

In this paper, I mainly focus on explaining data collection, metadata and data in detail. The Youtube project can be developed in various aspects, but the important point of this project is collecting Youtube data. Therefore, I will focus on data part only in here.

This data can help to understand many important questions. Which politicians use Youtube for what purposes? Are politicians Youtube contents different from their other media outlets such as Twitter and Facebook? How do politicians use their Youtube channel? How do politicians use emotions to affect audience? How do politicians use images? What is politicians campaign strategy?

## **Short literature review**

There is little literature in political science regarding politicians' Youtube channel and videos. Klotz (2010) studied how U.S. Senators use Youtube channels in 2008. Klotz (2010) states that only about 10 percent of Senators have their own Youtbe channel (Williams and Gulati 2007). However, in 2018 most of legislators have their own Youtube channels. Also Klotz argues that Senators' Youtbe channels are mostly composed of their advertisements but in 2018, they are mostly C-SPANs, hearing, news interview and messages (2010).

Borah et al. (2018) compared advertising in TV and Youtube, and showed that there are

no meaningful difference in their contents. However, they do not study all videos in Youtube but just focus on the advertisements.

Overall, it is very little known about legislators' Youtube channel contents. We do not even know much about what kind of videos are in there in descriptive level. Therefore, starting from descriptive level analysis to analyzing video contents, we need to explore Youtube data for political science.

## Data

### Collecting data

To begin with, I have total 484 legislators' Youtube channel id as of 2018 October. These include House and Senate members. This is complete list of channel that is available for legislators. It has 14265 urls and I have downloaded 5871 videos and 5155 captions that are about 300 GB in total. However, there is a problem in collecting this data. First, Youtube API does not allow users to grab all the videos from a channel (there is a limit). Second, the downloading process using Pytube package in Python does not allow me to download every videos from the given urls that I feed. Because of these two serious flaws, I restarted to collect full data from given legislators' Youtube channels.

The More robust way of collecting data is usingYoutube\_dl. Youtube\_dl provides some written Python package as well, so Youtube\_dl can work well with Python code as well. The strength of usingYoutube\_dl is that it allows users to grab and download all videos from a given channel. More importantly it does not get the data throughYoutube API, so the daily limit is not a problem withYoutube\_dl. However, there is also downside. The main problem is the written package or application is not easily malleable. For instance, I cannot just get the all ids ofYoutube videos from a channel. The functions are quite limited although there are many options for downloading videos becauseYoutube\_dl is specifically designed to download videos not archiving all the information. This is somewhat unsatisfying. For

instance, because `Youtube_dl` does not have option to just get all Youtube video ids of a channel, I can only collect them in process of downloading videos afterwards. However, `Youtube_dl` allows users to download the correct number of all videos and captions, so this is the best way to collect the data.

## Meta data: channel and video

The meta data will be two parts: Channel data and video data. Channel data have channel id, view count, subscriber count and video count. Video data have title, url, duration, author, thumbnail, video id, view count, likes and dislikes. To get channel meta data, I need to use `Youtube API` because `Youtube_dl` does not provide these information.

With this channel data, we can study who uses Youtbe more than other legislators for what purposes. The subscriber data and number of video data might be useful although the comparison among legislators using this information should be cautious. It is because some legislators have been in there office longer than others. However, this channel data can be useful meta data for further analysis by using video data.

With video meta data, we can investigate various questions. What types of videos are more popular than the others? What types of videos are uploaded in their youtube channel? More generally we can study how politicians use their Youtube channel different from their Twitter and Facebook. These questions are little explored in political science literature.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	title_main,caption	url	title_on_i	message	duration	author	length	thumbnail	viewcount	likes	dislikes			
2	011818C01	https://www.youtube.com/watch?v=Qz7Cfjyv7Co	002347 Doug Coll	00:00:47	0.00:47	Doug Coll	260	https://yt3.ggpht.com	96	2	0			
3	011727C01	https://www.youtube.com/watch?v=013J7TfCqyC	002347 Doug Coll	00:00:47	0.00:47	Doug Coll	237	https://yt3.ggpht.com	104	2	0			
4	011818C01	https://www.youtube.com/watch?v=01181818	00541 Doug Coll	00:04:41	0.04:41	Doug Coll	341	http://i.ytimg.com	69	0	0			
5	021318C01	https://www.youtube.com/watch?v=02131818	00440 Doug Coll	00:04:40	0.04:40	Doug Coll	280	http://i.ytimg.com	56	3	1			
6	022314Re1	https://www.youtube.com/watch?v=02231414	00232 Lauren Bu	00:04:32	0.04:32	Lauren Bu	152	http://i.ytimg.com	116	1	0			
7	022314Re1	https://www.youtube.com/watch?v=02231414	00232 Lauren Bu	00:04:32	0.04:32	Lauren Bu	179	http://i.ytimg.com	1018	1	0			
8	032013Re1	https://www.youtube.com/watch?v=03201313	01057 RepayFor	00:00:57	0.00:57	RepayFor	63	https://yt3.ggpht.com	125	1	0			
9	051818C01	https://www.youtube.com/watch?v=05181818	00137 Doug Coll	00:01:37	0.01:37	Doug Coll	97	https://yt3.ggpht.com	163	1	0			
10	051918C01	https://www.youtube.com/watch?v=05191818	00:02:56 Congress	00:02:56	0.02:56	Congress	26	https://yt3.ggpht.com	53	0	0			
11	062411Re1	https://www.youtube.com/watch?v=06241111	00125 Lauren Bu	00:01:25	0.01:25	Lauren Bu	85	https://yt3.ggpht.com	58	0	2			
12	062411Re1	https://www.youtube.com/watch?v=06241111	00125 Lauren Bu	00:01:25	0.01:25	Lauren Bu	196	https://yt3.ggpht.com	164	2	0			
13	062718C01	https://www.youtube.com/watch?v=06271818	00213 Doug Coll	00:02:13	0.02:13	Doug Coll	133	https://yt3.ggpht.com	157	3	0			
14	06272018	https://www.youtube.com/watch?v=06272018	00510 Nydia Vel	00:05:10	0.05:10	Nydia Vel	310	https://yt3.ggpht.com	51	3	2			
15	071118C01	https://www.youtube.com/watch?v=07111818	00440 Doug Coll	00:04:40	0.04:40	Doug Coll	280	http://i.ytimg.com	546	13	0			
16	072318C01	https://www.youtube.com/watch?v=07231818	00427 Doug Coll	00:04:27	0.04:27	Doug Coll	267	http://i.ytimg.com	133	17	0			
17	072318C01	https://www.youtube.com/watch?v=07231818	00427 Doug Coll	00:04:27	0.04:27	Doug Coll	277	http://i.ytimg.com	97	15	0			
18	072418C01	https://www.youtube.com/watch?v=07241818	00431 Doug Coll	00:04:31	0.04:31	Doug Coll	271	http://i.ytimg.com	54	8	0			
19	071618C01	https://www.youtube.com/watch?v=07161818	00552 Doug Coll	00:05:52	0.05:52	Doug Coll	352	http://i.ytimg.com	238	6	0			
20	090418C01	https://www.youtube.com/watch?v=09041818	00403 Doug Coll	00:04:03	0.04:03	Doug Coll	243	http://i.ytimg.com	174	11	0			
21	marvin18C01	https://www.youtube.com/watch?v=marvin1818	00418 Doug Coll	00:04:18	0.04:18	Doug Coll	248	http://i.ytimg.com	96	6	0			

Figure 1: Video metadata

	A	B	C	D	E
1	channel_id	num_vid	subscri	view_count	hidden_subscri
2	UCGdrLQbt1PYDTPsampx4t1A	112	0	17090	T
3	UCukn4jPGn91ghHDNAfb0ZEGQ	152	129	48174	F
4	UCLwrmtf_8AFIck3TyMs4Mlw	496	503	81683	F
5	UCUD9VGV4SSGWjGdbn37Ea2w	296	438	77585	F

Figure 2: Channel metadata

## Video

Using Youtube\_dl, I will download all videos in legislators' channels. Videos in Youtube are formatted in various types of video codecs. To make them uniform, I use FFmpeg to convert the videos into mkv file. Youtube\_dl is working well with FFmpeg software, so I can convert them. The downloaded videos are archived in each channel's folders.

All Files > Videos_master > ☆ videos		
Name	Updated	Size
Samoan language PSA advising youth to not use drugs.mp4	Feb 20, 2019 by Lim, Hyun Woo	46.9 MB
PSA advising against drug use.mp4	Feb 20, 2019 by Lim, Hyun Woo	24.1 MB
Talking with Todd Our spending problem in 2 minutes.mp4 [v2]	Feb 20, 2019 by Lim, Hyun Woo	11.5 MB
Talking with Todd Second Amendment Rights.mp4	Feb 20, 2019 by Lim, Hyun Woo	34 MB
Talking with Todd The REINS Act.mp4	Feb 20, 2019 by Lim, Hyun Woo	39 MB
Talking with Todd Why Im supporting a Balanced Budget Amend...	Feb 20, 2019 by Lim, Hyun Woo	28.6 MB

Figure 3: Data archive

## Transcript: caption

Downloaded captions of each video are useful. These text data can be used for machine learning to classify video types. There are several packages that provide transcription option, but Youtube already provide auto-generated captions for videos, so downloading these captions are good enough. However, not every video has this auto-generated captions unfortunately. In my preliminary data collecting shows that about 90 percent of videos has auto-generated captions. Here is the example of captions from former Rep. House Todd Young (current Senator) on Gun control. The Youtube video url is [Link](#):

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 .....

## **Comments**

Youtube video's comment is difficult to collect because, as far as I know, Youtube\_dl does not have function to download comments of each video. The only way is using Youtube API. To do so, I need complete list of urls, but I do not have the list because Youtube\_dl does not provide list of videos urls, but only can download all videos from channel. Therefore, the way to get all urls of these video is when I download the video, record the video id in separate document. After I finish downloading all videos, I will have the list of all videos that I have downloaded, and I can use this list to collect comments afterwards.

## **Findings, discussion and difficulties**

For the class, I use supervised learning method to classify videos. First, I hand-coded 100 videos of test set data to classify them with 9 types: cspan, hearing, message, mis, news, newsinterview, press, service and speech. This is my arbitrary coding method. From randomly selecting videos, in my observation, there are largely four types of videos in legislators' Youtube channel: C-SPAN, hearing, news interview of the legislator and message to Youtube viewers.



Figure 4: C-SPAN



Figure 5: Hearing



Figure 6: News interview



Figure 7: Message

In the future, we should use unsupervised learning to classify video type by using transcription data. When I used random forest model for supervised learning with 100 hand-coded training set and 100 test data, the accuracy was quite low. The accuracy rate with transcript was about 31 percent, and the accuracy rate with title of video was about 35 percent. The accuracy might be low because of following reasons: first, the training data and test data were too small. Second, the 9 types of video that I arbitrarily created are too many. Third, random forest model is not the best method to classify these kind of data.

Currently, I have a Window batch file that can collect all the data. The difficulty that I have now is incorporating command to send downloaded data from local drive to Box cloud data. As described, Youtube\_dl's code is difficult to work with python code that I wrote previously, so I am having difficulty to write a command for ftp for Box drive. As soon as I solve this problem, I can collect all data.

Another difficulty is collecting comments data. The only way is using Youtube API. However, I do not have complete list of Youtube videos (after I download all videos, I can

have the complete list of video ids) and unlimited quota for using Youtube API, I cannot collect comment data for now. I need to consult with Chris Lucas to figure out current problems.

## References

- Borah, Porismita, Erika Fowler, and Travis Nelson Ridout (2018). “Television vs. YouTube: political advertising in the 2012 presidential election”. In: *Journal of Information Technology & Politics* 15.3, pp. 230–244.
- Jeff Gulati, Girish J and Christine B Williams (2010). “Congressional candidates’ use of YouTube in 2008: Its frequency and rationale”. In: *Journal of Information Technology & Politics* 7.2-3, pp. 93–109.
- Klotz, Robert J (2010). “The sidetracked 2008 YouTube senate campaign”. In: *Journal of Information Technology & Politics* 7.2-3, pp. 110–123.
- Lev-On, Azi (2012). “YouTube usage in low-visibility political campaigns”. In: *Journal of Information Technology & Politics* 9.2, pp. 205–216.