

How does the Thai government conduct information operations on Twitter?

A Case Study of Jan.- Feb. 2020

Name: Sanittawan Tan
Date: March 19, 2020
Class: SOCI 40133
Computational Content Analysis, Winter 2020



Research Questions

- Exploratory project
- How does information operation work on the Thai social media landscape?
- Who are they?
- What are their goals?
- How are they different from the Chinese “50 cent” party?

Project Motivations

- Information Operations (IOs)
 - Military concept
 - Armistead (2010)
 - [Information Operation] is a formal attempt by the U.S. government to develop *a set of doctrinal approaches* for its military and diplomatic forces to *use and operationalize the power of information*. Per the original primary DOD policy on IO, the *target is the adversary decision makers*. Therefore, the *primacy of effort will be to coerce that person, or group of people, into doing or not doing a certain action* (Joint Publication 3-13 1998). To affect the adversary decision maker, IO attempts to use many different capabilities, such as deception, psychological operations, and electronic warfare, to shape and influence the information environment.

Project Motivations

- Various news sources reported IO activities, but no quantitative proof
- Early February 2020, Thai Twitter users noticed the trending of a suspicious hashtags “#รัฐบาลเตรียมพร้อม” (translation: #ThaiGovernmentisGettingReady)
 - Very similar user profiles
 - Very similar Tweets content

The screenshot shows a Twitter thread with two tweets. The first tweet is from a user named '1984 @newman1190' dated Feb 13, 2020. The text reads: 'ตัวอย่าง IO ก้าวๆ Copy ค่าพุ่มมาเลย 555555 #รัฐบาลเตรียมพร้อม'. The second tweet is from a user named 'Poppylove @Poppylove64434181' Replying to @WassanaNanum. The text reads: 'ใช้ครับ ถูเจ้าหน้าที่รัฐตั้งใจ ใช้ครับ ถูเจ้าหน้าที่รัฐตั้งใจ ทำมาก #รัฐบาลเตรียมพร้อม'. Both tweets have a red box around the hashtags '#รัฐบาลเตรียมพร้อม'.

The screenshot shows a Twitter search results page for the hashtag '#รัฐบาลเตรียมพร้อม'. It displays several tweets from different users, all containing the same or very similar text: 'ตามหกูดูต้นน้ำๆ IO คือ Account ไหนบ้าง พอยไปล่อเมือง #รัฐบาลเตรียมพร้อม ตอนที่ซื้อเม็ดเทียนหิน นี่จอกวั่นเลศดันภาระหนัก Report ให้คน Block ชั่วโมง เหมือนด่าวงไวย์จ้าว กส. ไปกดยอน รวมตัวได้ยกงั้น #รัฐบาล'. A red box highlights one such tweet from 'Wiroj @wirojlk'.

The screenshot shows a Twitter timeline for the user 'Wiroj @wirojlk'. It displays three tweets from this account, all containing the same text: 'เบคดีในการแสดง IO คือ เชิงป้อนเน้นๆ ในเว็บต่อรัฐบาล ที่คนร่วมกากา ล้างของ account ไหน ที่ซื้อเมปล็อก ให้รูปป้อน Follow คนแม่ไม่เกิน นานนั้นที่ต้อง คลิก ตอนได้ ให้เข้าไปลองใน Timeline ล้านเมล์หรือรัฐบาล เชิญร่องทัพ กี Block โลก แผนผังเข้าไปเช็ค คนที่เข้า Follow แล้ว Block มี'. The last two tweets also mention 'Report' and 'Block'.

Research Goals and Tasks – Overview

Steps:

1. Data Collection
2. Build a classifier to distinguish between suspected IO accounts and other users
3. Analyze the suspected IO accounts
4. Explore the content of the IO accounts

Ultimate Goal: Use this classifier with other historical tweets to make inference about how IO operates (future project)

Step 1: Data Collection

- Twitter only allows us to collect and search tweets for the past 7 days
 - Use [Twitter Scraper](#) package to collect historical Twitter tweets
 - Gather all tweets with the hashtag “#รัฐบาลเตรียมพร้อม” (translation: #ThaiGovernmentisGettingReady) in Thai from January 28, 2020 to February 20, 2020
 - 3,432 tweets gathered
- Deal with missing information on users (number of tweets, date joined etc.) with Twitter REST API

screen_name_tweets	fullname	text	timestamp	following	followers	num_tweets	user_id_users	date_joined	has_media
trueman98587845	trueman	สู้ นศรับ เป็นกำลังใจให้ เจ้าหน้าที่ทุกคน	2020-01-28 13:2	30	4	1032	1220152441666	5:13 PM - 22 Jan	FAL
KI36bsxg8ozf6bg	ชัย สุขากย์สนายา	ใช่ #รัฐบาลเตรียมพร้อม	2020-02-03 15:5	36	22	2673	1201755584225	10:50 PM - 2 Dec	FAL
		ช่วยป้อน ไม่จริง ไม่ควรแชร์ต่อ #ไว้สักโคน #ไว้รีบอู่น้ำ #รัฐบาลเตรียมพร้อม #สภาพอากาศมีเม็ด							
daengbile436	แดง ไบเล่	#อย่าทิ้งคนไทยhttps://twitter.com/daengl	2020-01-29 3:46	38	5	1234	1219487869049	9:12 PM - 20 Jan	FAL
trueman98587845	trueman	กรณะหนึ่งในเพื่อนคนไทย ขอเป็นกำลังใจ	2020-01-31 3:22	30	4	1032	1220152441666	5:13 PM - 22 Jan	FAL
		ขอร้องคุณช่วยเผยแพร่ถือย่ามีสติ อวยพรด้วย							
thaiprong	รวมใจไทยปrongด	#อย่าทิ้งคนไทยhttps://twitter.com/suchar	2020-01-28 8:28	304	99	78	1209024400651	12:14 AM - 23 Dec	FAL

Step 2: Build a classifier

2 Main Challenges

1. Lack of “ground truth”

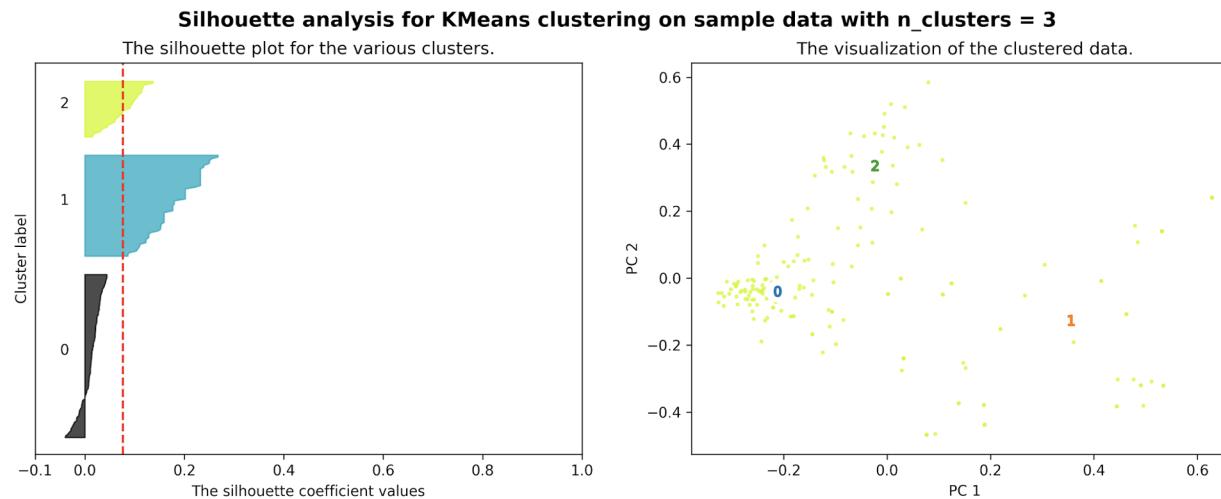
- a. The data set has no label. No “official” information on which accounts are definitely IOs.
- b. Solution: Manually label a sample of the data (300 observations)

2. The Thai language

- a. The language makes it a bit challenging to tokenize words using the general method
 - i. “ข่าวปลอม” (fake news) is supposed to be “ข่าว” (news) and “ปลอม” (fake). Without specialized package, it becomes gibberish e.g. “ຂ”, “ງ”, “ປ”, “ລອມ”
- b. Solution: Use specific Python packages to tokenize words and bi-grams after tokenization

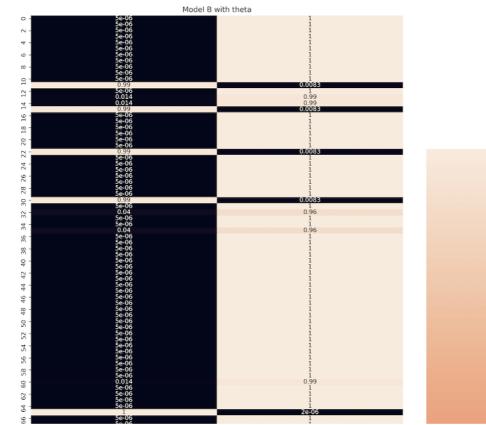
Step 2: Build a classifier

- Use k-means clustering ($k = 3$) to cluster the data based on tweets



Step 2: Build a classifier

- Create a training set of 300 observations drawn randomly
- 3 annotators: Recruit 2 more native Thai language speakers to label the accounts
 - Similar level of education, different fields of study
 - Both follow Thai politics closely
- Inter-coder agreement
 - Fleiss' kappa and Cohen's kappa ~93.8%
- Breakdown →



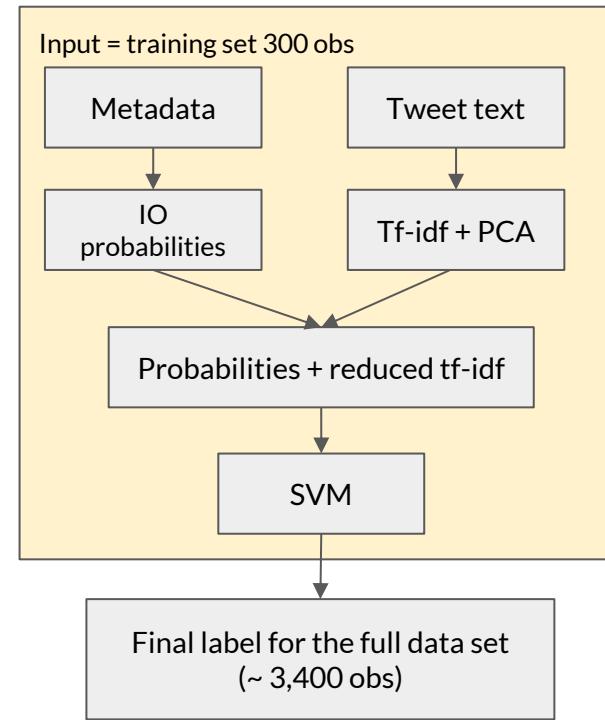
Model B with Theta to resolve labeling conflict

	Training Set	Test Set
# IO	75 (~40%)	31 (~40%)
# not IO	118 (~60%)	51 (~60%)

Step 2: Build a classifier

- Idea: mixed classifier
 - Generate probabilities of being an IO accounts from metadata e.g. number of tweets, following, followers, number of days on Twitter etc.
 - XGBoost 5-fold CV accuracy rate 97% (train set)
 - Test set accuracy rate 98%
 - Create tf-idf matrix of tweet texts and use PCA to reduce dimension to 3 dimension
 - Append predicted probabilities to this matrix
 - Train and use SVM to classify the full data set
 - 5-fold CV accuracy rate 98%
 - Test set accuracy rate 100%
- Result:
 - 375 IO posts out of 3,405 posts
 - Suspected IO accounts = ~11% of the data set

Model Building flowchart



Step 3: Explore suspected users

The image shows a Twitter search results page with two main user profiles highlighted by red boxes and connected by red arrows.

User 1: trueman (Left Column)

- Profile:** A cat icon with "404".
- Followers:** weloverta, @weloverta (Follow), ชานมายาหวาน (Follow), กองพัทภาคที่ 4 (@Army4PR) (Follow), army3news.pr (Follow), First Army Area (@1st_Army_Area) (Follow), Mouseymouse (@mousey2355) (Follow), Somchai PuckFavit (@PravitR) (Follow), PDLS (@MR_PDLS) (Follow).
- Tweets:** 1. กรณีที่รัฐบาลจัดการเรื่องน้ำดื่มในประเทศไทย ให้ห้ามสูบบุหรี่ในสถานที่สาธารณะ ทำให้คนต้องหันมาซื้อน้ำดื่มจากตลาด黑市 2. กรณีที่รัฐบาลจัดการเรื่องน้ำดื่มในประเทศไทย ให้ห้ามสูบบุหรี่ในสถานที่สาธารณะ ทำให้คนต้องหันมาซื้อน้ำดื่มจากตลาด黑市 3. กรณีที่รัฐบาลจัดการเรื่องน้ำดื่มในประเทศไทย ให้ห้ามสูบบุหรี่ในสถานที่สาธารณะ ทำให้คนต้องหันมาซื้อน้ำดื่มจากตลาด黑市
- Media:** A video thumbnail showing a group of men in suits walking outdoors.
- Likes:** 0:23 | 212.8K views

User 2: Chai Sukkai Sabajaijai (Right Column)

- Profile:** A cat icon.
- Followers:** ชัย สุขากาญจน์ (Follow), Matichon Online (@MatichonOnline) (Follow), ใจดีชัย (@lodchai40277) (Follow), AFNC Thailand (@AFNCThailand) (Follow), กองพัทภาคที่ 4 (@Army4PR) (Follow), army3news.pr (Follow), First Army Area (@1st_Army_Area) (Follow).
- Tweets:** 1. หมุดเตือน ศือพุดไปจริง พูดตรงกันข้ามกับความจริง เพื่อให้เช่าใจผิด 2. หมุดค่าหมาย ศือพุดค่าระคายๆ ค่าหมายโน่น ไม่ naïve ไม่เป็นค่าถูกต้อง 3. หมุดสองเส้น ศือพุดดูยังไงให้คนอื่นแตกดกัน

User 1: trueman

User 2: Chai Sukkai Sabajaijai

Step 3: Explore suspected users

- In general, suspected IO accounts have not been on Twitter for a long time, have low number of following and followers and relatively low number of Tweets

	following	followers	num_tweets	num_likes	id	days_on_twitter
count	375.00	375.00	375.00	375.00	3.750000e+02	375.00
mean	106.69	8.71	709.43	2.69	1.202177e+18	74.77
std	102.21	18.34	774.64	23.84	1.185704e+17	239.21
min	2.00	0.00	1.00	0.00	4.515064e+08	13.00
25%	34.00	2.00	147.00	0.00	1.212928e+18	14.00
50%	53.00	4.00	493.00	0.00	1.220152e+18	35.00
75%	190.00	5.00	1032.00	1.00	1.227765e+18	55.00
max	811.00	133.00	4606.00	420.00	1.228032e+18	2979.00

Step 4: Explore the content

- Utilize topic modeling to explore the content of the suspected IO posts
- Unigram approach
- The result suggests that the content is not diverse at all
 - A lot of overlapping words across all topics
- Posts focus on the Coronavirus outbreak and the government
 - Words are such as “getting ready”, “Wuhan”, “government”, “virus”, “media”, “consciousness”

	Topic_0	Topic_1	Topic_2	Topic_3	Topic_4
0	พร้อม	เดรีym	รัฐบาล	พร้อม	ขอ
1	รัฐบาล	ไวรัส	เดรีym	เดรีym	ให้
2	เดรีym	รัฐบาล	พร้อม	รัฐบาล	ท่าน
3	ให้	พร้อม	ไวรัส	ไม่	คน
4	ไวรัส	ให้	ช่าว	ไวรัส	ทุก
5	ที่	โควิด	ให้	มี	เดรีym
6	อย่าง	ท่าน	ที่	อย่า	และ
7	คน	ทุก	ครับ	อย่าง	ไวรัส
8	มี	มี	มี ตื่นตระหนก	ที่	
9	ได้	คน	อย่า	สติ	พร้อม
10	ทำ	สภาพ	ปลอม	สภาพ	มี
11	สื่อ ตื่นตระหนก		สติ	สื่อ	รัฐบาล
12	อยู่บ้าน	สื่อ	ได้	อยู่บ้าน	กำลังใจ

Step 4: Explore the content

- Utilize topic modeling to explore the content of the suspected IO posts
- Unigram approach
- The result suggests that the content is not diverse at all
 - A lot of overlapping words across all topics
- Posts focus on the Coronavirus outbreak and the government
 - Words are such as “getting ready”, “Wuhan”, “government”, “virus”, “media”, “consciousness”

	Topic_0	Topic_1	Topic_2	Topic_3	Topic_4
0	พร้อม	เดรีym	รัฐบาล	พร้อม	ขอ
1	รัฐบาล	ไวรัส	เดรีym	เดรีym	ให้
2	เดรีym	รัฐบาล	พร้อม	รัฐบาล	ท่าน
3	ให้	พร้อม	ไวรัส	ไม่	คน
4	ไวรัส	ให้	ช่าว	ไวรัส	ทุก
5	ที่	โควิด	ให้	มี	เดรีym
6	อย่าง	ท่าน	ที่	อย่า	และ
7	คน	ทุก	ครับ	อย่าง	ไวรัส
8	มี	มี	มี ตื่นตระหนก	ที่	
9	ได้	คน	อย่า	สติ	พร้อม
10	ทำ	สภาพ	ปลอม	สภาพ	มี
11	สื่อ ตื่นตระหนก		สติ	สื่อ	รัฐบาล
12	อยู่บ้าน	สื่อ	ได้	อยู่บ้าน	กำลังใจ

Step 4: Explore the content

- bigram approach
- Again, the result suggests that the content is not diverse at all
 - A lot of overlapping words across all topics
- Posts focus on the Coronavirus outbreak and the government
 - Words are such as “getting ready”, “Wuhan”, “government”, “virus”, “media”, “consciousness”

	Topic_0	Topic_1	Topic_2	Topic_3	Topic_4
0	รัฐบาล เตรียมพร้อม	รัฐบาล เตรียมพร้อม	รัฐบาล เตรียมพร้อม	รัฐบาล เตรียมพร้อม	รัฐบาล เตรียมพร้อม
1	มีสติ อย่า	ไวรัส โควิด	โควิด	ช่วง ปลอม	กำลังใจ ให้
2	เสพ สื่อ	โควิด	ไวรัส โควิด	ไวรัส โควิด	โควิด
3	อุ่น ยั่น	ช่วง ปลอม	อย่า ตื่นตระหนก	โควิด	อุ่น ยั่น
4	นา ไวรัส พยายาม ประชาสัมพันธ์	เสพ สื่อ	ไวรัส โควิด	โควิด	ยั่น รัฐบาล
5	อัน รัฐบาล	สื่อ อย่าง	อย่าง มีสติ	มีสติ อย่า	โควิด
6	อย่าง มีสติ	โควิด	โควิด	ไวรัส อุ่น	เตรียมพร้อม เสพ
7	โควิด	ครวัต รัฐบาล	สื่อ อย่าง	สื่อ อย่าง	ไวรัส โควิด
8	โควิด	ไวรัส อุ่น	ไวรัส อุ่น	อย่า ตื่นตระหนก	ไวรัส อุ่น
9	สื่อ อย่าง	ประชาสัมพันธ์ ให้	เตรียมพร้อม เสพ	เตรียมพร้อม เสพ	นา ไวรัส
10	เตรียมพร้อม เสพ	อย่าง มีสติ	อุ่น ยั่น	ยั่น รัฐบาล	สื่อ อย่าง
11	ไวรัส โควิด	อัน รัฐบาล	มีสติ อย่า	อุ่น ยั่น	อย่าง มีสติ
12	ไวรัส อุ่น	นา ไวรัส	ยั่น รัฐบาล	เสพ สื่อ	พวก ท่าน
13	อย่า ตื่นตระหนก	อย่า ตื่นตระหนก	นา ไวรัส	อย่าง มีสติ	เสพ สื่อ
14	พวก ท่าน	อุ่น ยั่น	กำลังใจ ให้	นา ไวรัส	มีสติ อย่า
15	กำลังใจ ให้	เตรียมพร้อม เสพ	พวก ท่าน	fakenews ช่าว	อย่า ตื่นตระหนก
16	ช่วง ปลอม	เสพ สื่อ	สายพันธุ์ ใหม่	ใน การ	ที่ ก้าว
17	เป็น กำลังใจ	มีสติ อย่า	เป็น กำลังใจ ตื่นตระหนก fakenews	เป็น กำลังใจ	เป็น กำลังใจ
18	ให้ พวก	ความตื่นตัว ใน	ใหม่ 2019	วิเคราะห์ ด้วย	ขอ เป็น
19	ประชาสัมพันธ์ ให้	รัฐ พยายาม	ที่ พึ่ง ของ	ไม่ จริง	หน้า อยู่

Step 4: Explore the content

- Result: exact same wordings by different users and time stamp
- Utilize similarities of documents to evaluate posts (similarities > 0.5)

	screen_name	fullname	cleaned_text	timestamp	is_IO
70	tawannak32	ทวารหนัก	ข่าวปลอมครับ ในเรื่องไวรัสโควิด ไม่ข่าวปลอมจ้าน...	2020-01-29 12:37:52	1
94	tawannak32	ทวารหนัก	ข่าวปลอมครับ ใช้วิจารณญาณในการวิเคราะห์ด้วยนั่...	2020-01-30 13:34:49	1
99	sakonkob2	สกุลวิชิต	ข่าวปลอมครับ ใช้วิจารณญาณในการวิเคราะห์ด้วยนั่...	2020-01-30 13:33:12	1
100	sakonkob2	สกุลวิชิต	ข่าวปลอมครับ ใช้วิจารณญาณในการวิเคราะห์ด้วยนั่...	2020-01-30 13:32:51	1

Concluding Points

1. The content of the suspected IO accounts is very coherent
 - a. The hashtag (#รัฐบาลเตรียมพร้อม; #ThaiGovernmentisGettingReady) seems to be invented to cope with the covid-19 outbreak
 - b. It looks like an attempt for the government to set an agenda/propagandize their work

Concluding Points

1. The content of the suspected IO accounts is very coherent
 - a. The hashtag (#รัฐบาลเตรียมพร้อม; #ThaiGovernmentisGettingReady) seems to be invented to cope with the covid-19 outbreak
 - b. It looks like an attempt for the government to set an agenda/propagandize their work
2. One of their strategies is to create multiple fake accounts to send the same messages
 - a. Various accounts with the same exact posts/wordings but different timestamps
 - b. Content is also similar

Next Steps

1. Use the classifier or the same method on more salient political events
 - a. Collect tweets during important political events ex. the 2019 Thai general election
 - b. Assess the number of IO activities and accounts
 - c. Question: Do we see spikes in IO activities correlated with important political events?

Next Steps

1. Use the classifier or the same method on more salient political events
 - a. Collect tweets during important political events e.g. the 2019 Thai general election
 - b. Assess the number of IO activities and accounts
 - c. Question: Do we see spikes in IO activities correlated with important political events?
2. Once sufficient IO accounts have been gathered, ...
 - a. Build a network of IO accounts on Twitter
 - b. Study the content of their posts
 - c. Question: Are they similar to the 50 cent party?

Any suggestions are most welcomed! 😊