

자연어처리 및 정보검색, 정재은 교수님

TOPIC MODELING BY AMERICAN NEWS

20190615

HOLA TEAM

Irene Llop Escrich

허수민

김희선

박채린

이은주

Laura Nienajadlo

INDEX

1. 프로젝트 소개
2. 주간 별 진행 과정
3. 결과 분석 및 결론

PROJECT DESCRIPTION

주제 : 정치적 성향에 따른 미국 언론사 간의 **사용 어휘 차이 분석**



진보

VS



보수

PROJECT DESCRIPTION



TOOLS

LANGUAGE

- Python 3



IDE

- Jupiter Notebook



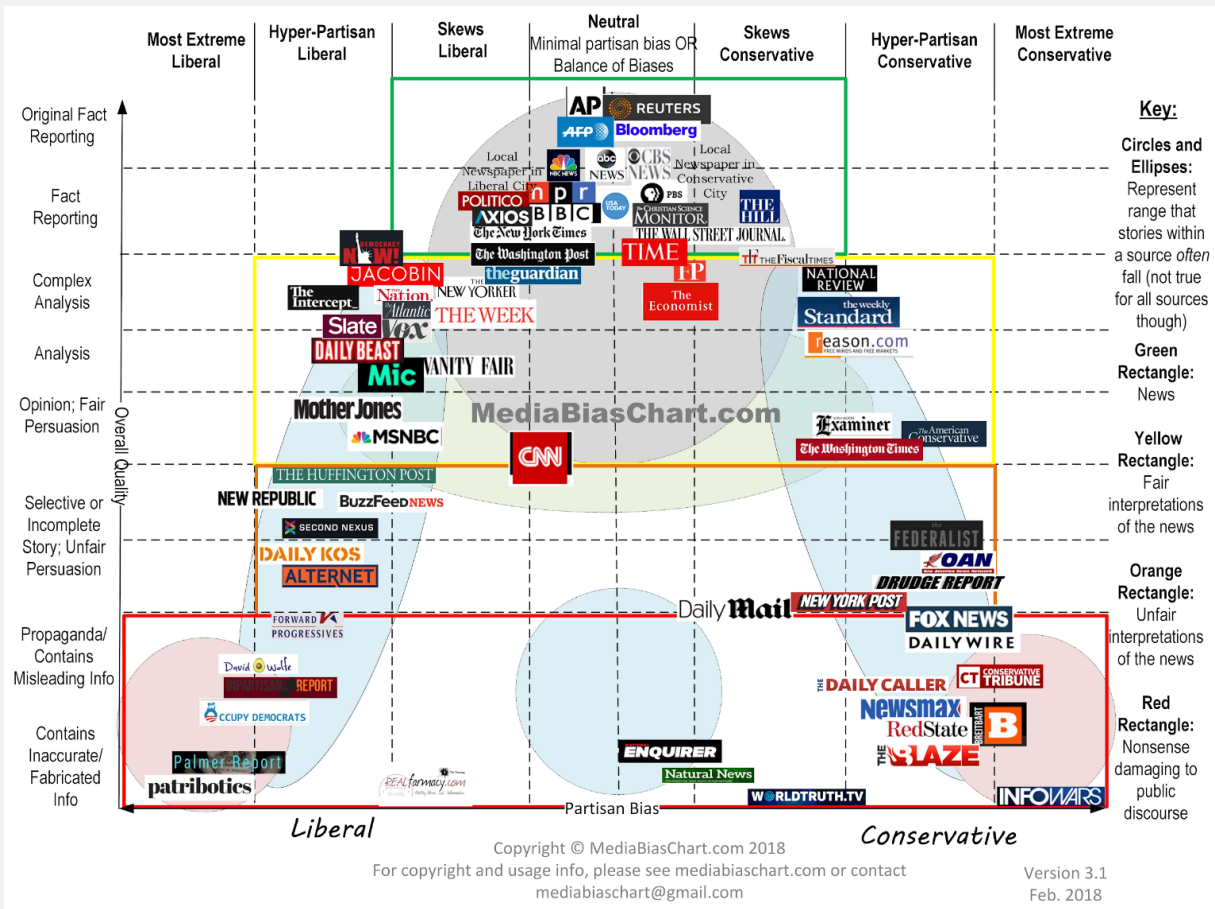
Collaboration Tool

- Github
- Google Drive



DATASET

데이터 : 미국의 다양한 언론사 별 기사 32000개 (진보 16000 + 보수 16000)



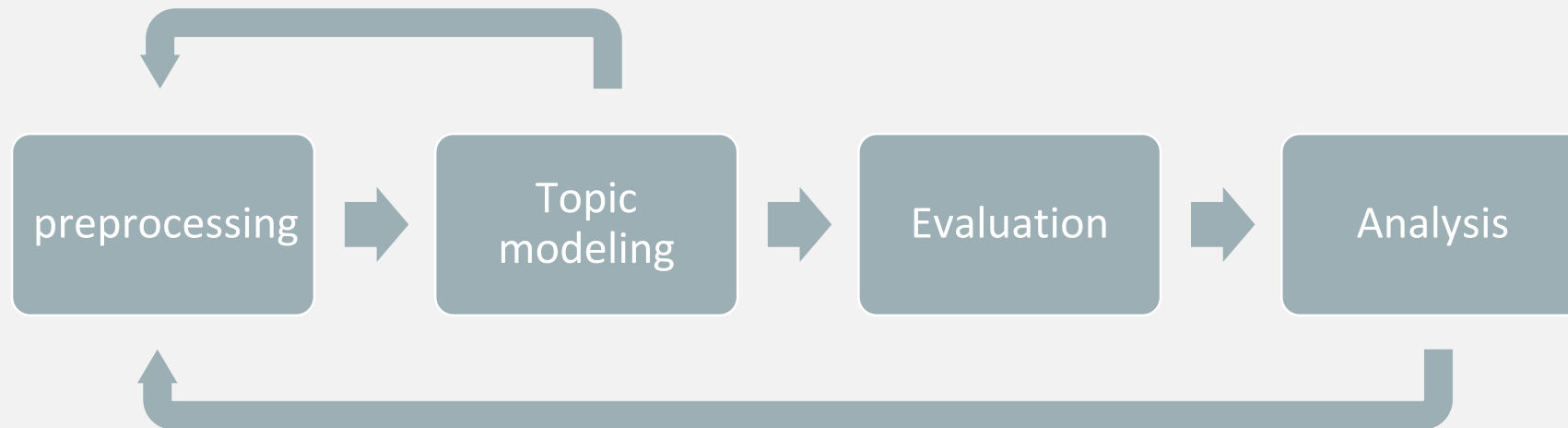
- liberal (Atlantic, BuzzFeed News, Vox)
- conservative (Breitbart, New York Post, National Review, Fox News)

(출처 : Dataset: <https://www.kaggle.com/snapcrack/all-the-news>)

WEEKLY PROJECT PROCESS

| | 1st Week | 2nd Week | 3rd Week | 4th Week | 5th Week |
|----------------|----------|----------|----------|----------|----------|
| Proposal | | | | | |
| Pre-processing | | | | | |
| Topic modeling | | | | | |
| Evaluation | | | | | |
| Analysis | | | | | |
| Final report | | | | | |

WEEKLY PROJECT PROCESS



PREPROCESSING

- 1) 언론사 성향을 기반으로 진보, 보수 기사 분류
- 2) 16000개의 기사 선정
- 3) TEXT OPERATION – NLTK 사용
- 4) 중간 테스트 결과를 반영한 추가 단어 삭제

PREPROCESSING

| | A | B | C | D | E | F | G | H | I | J |
|---|---|-------|------------|-------------|-------------|------------|------|-------|-----|-------------|
| 1 | | id | title | publication | author | date | year | month | url | content |
| 2 | 0 | 17283 | House Ref | New York | Carl Hulse | 2016-12-31 | 2016 | 12 | | WASHING |
| 3 | 1 | 17284 | Rift Betwe | New York | Benjamin I | 2017-06-19 | 2017 | 6 | | After the b |
| 4 | 2 | 17285 | Tyrus Wor | New York | Margalit Fe | 2017-01-06 | 2017 | 1 | | When Walk |
| 5 | 3 | 17286 | Among De | New York | William Mc | 2017-04-10 | 2017 | 4 | | Death may |
| 6 | 4 | 17287 | Kim Jong-i | New York | Choe Sang | 2017-01-02 | 2017 | 1 | | SEOUL, So |
| 7 | 5 | 17288 | Sick With | New York | Sewell Cha | 2017-01-02 | 2017 | 1 | | LONDON |

- 1) 소문자 변경
- 2) 구두점, 인용부호, BAR 제거
- 3) TOKENIZE
- 4) STOPWORDS 제거
- 5) 어간 추출

TOPIC MODELING

- 1) 전처리 된 데이터를 불러와 BAG OF WORDS로 만듦
- 2) 해당 데이터를 이용해 TF-IDF VECTORIZATION
- 3) GENSIM의 LDA 모델 생성 함수 LDAMULTICORE 이용
 - LDA 모델 생성, topic 15개로 나눔

TROUBLE SHOOTING

- 전처리

- 무의미한 TOPIC들이 추출되고, 공통적으로 많이 등장하는 단어로는 유의미한 결과를 내기 어려움

>> 진보 - 보수 언론에서 공통으로 언급된 단어 중 빈도 차이가 0.000001 이하인 단어 삭제 (ex. Trump)

```
./LDAmode1/liberal_15.model  
Topic: 0 Word: 0.002*"trump" +  
Topic: 1 Word: 0.001*"trump" +  
Topic: 2 Word: 0.001*"trump" +  
Topic: 3 Word: 0.001*"trump" +  
Topic: 4 Word: 0.001*"trump" +  
Topic: 5 Word: 0.001*"trump" +  
Topic: 6 Word: 0.001*"opioid" +  
Topic: 7 Word: 0.007*" " + 0.004  
Topic: 8 Word: 0.003*"trump" +  
Topic: 9 Word: 0.001*"trump" +  
Topic: 10 Word: 0.001*"trump" +  
Topic: 11 Word: 0.001*"trump" +  
Topic: 12 Word: 0.002*"trump" +  
Topic: 13 Word: 0.003*"zika" +  
Topic: 14 Word: 0.001*"trump" +
```

TROUBLE SHOOTING

>> AUTHOR, PUBLICATION 칼럼을 참고해서 기사에 포함된 기자, 언론사 이름 및 SNS계정 삭제

everyon retir fund georg soro netflix astronom ratio 340 simpl term mean stock price predic massiv continu growth profit last thing netflix need right shed custom netflix amaz run clear leader stream even movi select dodgi continu achiev new height origin program leav us one tremend question piss mani subscrib potenti subscrib temporari glow gain good liber worth pain liber simpli never learn money latter evid right daili dot label anyon cancel netflix white supremacist avow marxist left invent new form econom let call capit goe like subscrib netflix white supremacist drink starbuck hate muslim good luck lefti conserv already know vote dollar unlik republican actual job also unlik liber quit social media cri harass conserv come back brand two week later netflix win back custom current alien radic idea entertain show vilifi peopl race shock know danger avail via amazon hardcov kindl edit ye milo read audiobook version follow milo yiannopoulo nero facebook instagram snapchat hear everi friday milo yiannopoulo show write milo milobreitbart com wikileak threaten sue cnn defam network air clip cnn counterterror analyst phil mudd accus wikileak founder julian assang pedophil tweet offici wikileak task forc account support account organis tweet read cnn air cia fals call assang pedophil cnn air cia fals call assang pedophil http plot http wikileak task forc wltaskforc januari 4 2017 clip attach tweet show phil mudd cnn new day program discuss recent fox news interview julian assang mudd accus fox news sean hanniti skirt bottom line relat question russian interfer us elect despit assang openli state wikileak sourc dnc document way relat russia mudd state thought effort wikileak protect pedophil live ecuadorian embassi london claim julian assang credibl respons offici wikileak twitter account retweet task forc tweet ad quot issu instruct sue cnn defam clarifi unless within 48h air one hour expos plot issu instruct sue cnn defamationhttp unless within 48h air one hour expos plot wikileak wikileak januari 4 2017 cnn new day twitter account sinc delet tweet wikileak task forc cite appear remov segment video mudd call assang pedophil stori tweet origin link also tweet statement say cnn evid support accus

보수 언론사 Breitbart 기자의 SNS 홍보문

TOPIC MODELING

- 여러 개의 토픽수(10,15,20,25)로 모델링을 진행했을 때 단어의 분포가 어떻게 변하는 지 분석

| 2019-6-8 | saved_liberal.txt 15 | 2019-6-8 | saved_conserv.txt 15 |
|-----------|---------------------------------------|-----------|--|
| Topic: 0 | Word: 0.001*"erdogan" + 0.001*"music" | Topic: 0 | Word: 0.002*"islam" + 0.001*"obama" + |
| Topic: 1 | Word: 0.001*"women" + 0.001*"prison" | Topic: 1 | Word: 0.003*"percent" + 0.002*"poll" + |
| Topic: 2 | Word: 0.001*"weiner" + 0.001*"women" | Topic: 2 | Word: 0.002*"caddel" + 0.001*"discuss" |
| Topic: 3 | Word: 0.001*"russia" + 0.001*"obama" | Topic: 3 | Word: 0.002*"hillari" + 0.002*"cruz" + |
| Topic: 4 | Word: 0.001*"russia" + 0.001*"eu" + | Topic: 4 | Word: 0.001*"scarborough" + 0.001*"hil |
| Topic: 5 | Word: 0.001*"marijuana" + 0.001*"vic | Topic: 5 | Word: 0.001*"attack" + 0.001*"student" |
| Topic: 6 | Word: 0.000*"obama" + 0.000*"republ | Topic: 6 | Word: 0.001*"cruz" + 0.001*"prptest" + |
| Topic: 7 | Word: 0.001*"republican" + 0.001*"va | Topic: 7 | Word: 0.002*"republican" + 0.002*"ryan |
| Topic: 8 | Word: 0.002*"republican" + 0.002*"sa | Topic: 8 | Word: 0.001*"migrant" + 0.001*"border" |
| Topic: 9 | Word: 0.003*"zika" + 0.002*"mosquito | Topic: 9 | Word: 0.002*"cruz" + 0.002*"rubio" + 0 |
| Topic: 10 | Word: 0.001*"pipelin" + 0.000*"trib | Topic: 10 | Word: 0.002*"zika" + 0.001*"percent" |
| Topic: 11 | Word: 0.001*"film" + 0.001*"movi" | Topic: 11 | Word: 0.001*"attack" + 0.001*"migrant" |
| Topic: 12 | Word: 0.001*"tax" + 0.000*"student" | Topic: 12 | Word: 0.001*"protest" + 0.001*"senat" |
| Topic: 13 | Word: 0.001*"senat" + 0.001*"repub | Topic: 13 | Word: 0.001*"refuge" + 0.001*"palesti |
| Topic: 14 | Word: 0.001*"nune" + 0.001*"women" | Topic: 14 | Word: 0.001*"attack" + 0.001*"state" |

Zika virus outbreak(2016)

- ‘버락 오바마 미 11억 달러 지카 대책 예산, 시민단체 가족계획(Planned Parenthood)에 대한 예산 지원 조항 때문에 또 무산’

- **Conservative**

Topic N = 10

Word: **0.001*"zika"** + 0.001*"2017" + 0.001*"democrat" + 0.001*"news" + 0.001*"state" + **0.001*"obama"** + 0.001*"campaign" + 0.001*"attack" + 0.001*"elect" + 0.001*"white"

Topic N = 15

Topic: 10 Word: **0.002*"zika"** + 0.001*"percent" + **0.001*"women"** + 0.001*"state" + 0.001*"poll" + 0.001*"broaddrick" + 0.001*"attack" + 0.001*"news" + 0.001*"vote" + **0.001*"obama"**

Topic N = 20

Topic: 2 Word: **0.002*"zika"** + **0.001*"obama"** + 0.001*"state" + 0.001*"lynch" + 0.001*"fbi" + 0.001*"campaign" + **0.001*"health"** + **0.001*"2016"** + 0.001*"news" + 0.001*"republican"

Topic N = 25

Topic: 4 Word: **0.003*"zika"** + 0.002*"cartel" + 0.002*"border" + 0.001*"migrant" + 0.001*"mexican" + 0.001*"abort" + **0.001*"mosquito"** + 0.001*"cair" + **0.001*"parenthood"** + 0.001*"agent"

Standing Rock protest: against the Dakota pipeline(2016)

- ‘노스다코타 주 원주민 보호구역을 통과해 논란이 일었던 ‘다코타 액세스 송유관(pipeline) 건설’이 수많은 원주민(tribe)과 연대자들의 거센 반발에 중단되었다.’

- **Liberal**

Topic N = 10

None

Topic N = 15

Topic: 10 Word: 0.001*"pipelin" + 0.000*"tribe" + 0.000*"protest" + 0.000*"state" + 0.000*"women" + 0.000*"refuge" + 0.000*"yahoo" + 0.000*"court" + 0.000*"countri" + 0.000*"driver"

Topic N = 20

Topic: 6 Word: 0.001*"russian" + 0.001*"russia" + 0.001*"intellig" + 0.001*"attack" + 0.001*"pipelin" + 0.001*"assad" + 0.001*"emoji" + 0.001*"syria" + 0.001*"news" + 0.001*"administr"

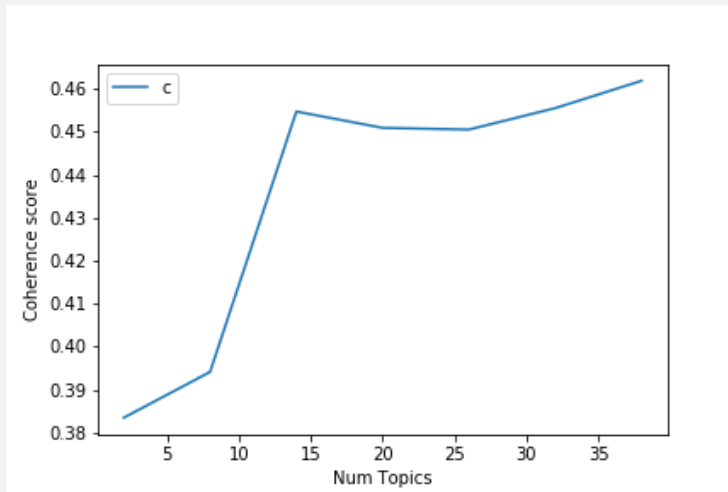
Topic N = 25

Topic: 18 Word: 0.001*"pipelin" + 0.001*"republican" + 0.001*"administr" + 0.001*"daca" + 0.001*"food" + 0.001*"state" + 0.001*"senat" + 0.001*"democrat" + 0.001*"sisi" + 0.000*"order"

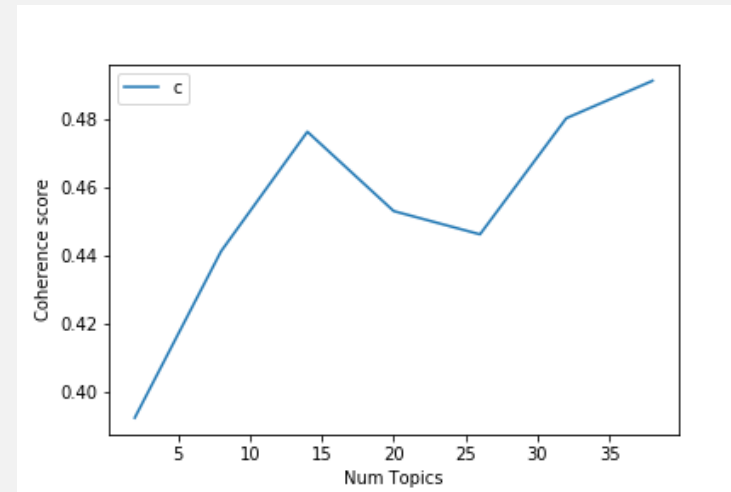
* DACA(미성년 입국자 추방 유예, 미성년자의 국외추방을 유예하는 민법 제도)

EVALUATION (TOPIC COHERENCE)

Liberal



Conservative

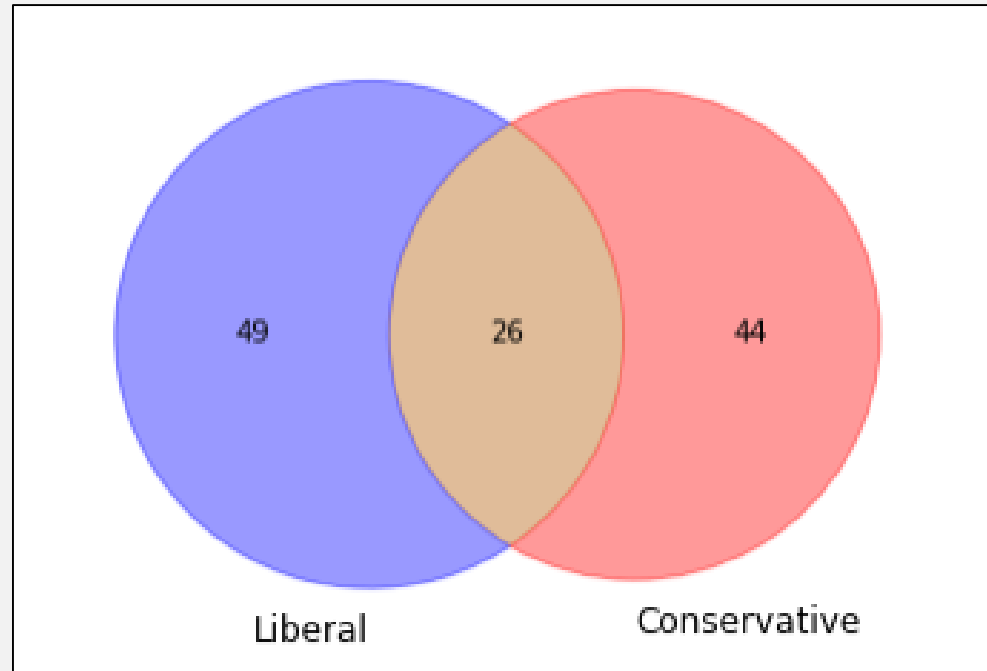


- 대체적으로 N이 높아질수록 Topic의 정확도가 높아졌지만, 한 단어가 여러 토픽에서 동시에 등장하는 경우도 증가하며 정확도 감소
- Topic num이 15일 때 Topic Coherence score 가장 높게 나타남

ANALYSIS

- 어휘 차이 분석
 - 1) 각 진영의 모델에서 topic 마다 가중치가 가장 높은 단어 10개를 뽑음
→ 총 150개의 단어에서 중복단어 제거 → liberty 75개, conservative 71개 사용
 - 2) 어휘 비교를 위해 벤다이어그램 생성

ANALYSIS



<Venn Diagram - Liberal & Conservative>

ANALYSIS

LIBERAL ONLY: CHINA KOREA EU GOVERN BLACK EMISS EMOJI US ALBUM TAX MUSIC MARIJUANA TRANSGEND UPDAT CAR CARBON COAL PREGNANT YAHOO TRACK INTELLIG RUSSIA TILLERSON ASSAD TESLA HEALTH VACCIN MUSK ARCHIV INFECT SEASON DRIVER THRONE WATER READER POKÉMON FIRE PIPELIN TRIBE GUZMÁN MOVI SCIENC LOCHT SCHOOL CLIMAT SODA NUNE PRISON MOSQUITO

BOTH: TURKEY SENAT WHITE VOTE FILM SANDER STATE REPUBLICAN WOMEN ATTACK COURT ERDOGAN STUDENT CAMPAIGN AMERICAN ZIKA DEMOCRAT PROTEST NEWS ELECT RUSSIAN OBAMA VIDEO COUNTRI WEINER REFUGE

CONSERVATIVE ONLY: RUBIO ABEDIN TOLD BILL NEW LIVE GERMANI ISRAEL MEXICAN 2016 NFL GUN LAW RYAN MIGRANT TURKISH MEDIA FBI CARTEL ILLEG TWEET IOWA ISLAM PALESTINIAN CRIME MUSLIM BROADDRICK HILLARI 2017 CALL PERCENT DISCUSS PATRIOT ISRA CARSON BORDER GINGRICH SUPPORT CADDEL SCARBOROUGH KAIN JERUSALEM POLL CRUZ

1. 나라/분쟁지역

- Liberal - China, Korea, EU, Russia (분쟁이 적은 나라)
- Conservative - Israel, Islam, Palestinian, Muslim, Isra
- Both - Turkey

LIBERAL ONLY: CHINA KOREA EU GOVERN BLACK EMISS EMOJI US ALBUM TAX MUSIC MARIJUANA TRANSGEND UPDAT CAR CARBON COAL PREGNANT YAHOO TRACK INTELLIG RUSSIA TILLERSON ASSAD TESLA HEALTH VACCIN MUSK ARCHIV INFECT SEASON DRIVER THRONE WATER READER POKÉMON FIRE PIPELIN TRIBE GUZMÁN MOVI SCIENC LOCHT SCHOOL CLIMAT SODA NUNE PRISON MOSQUITO

BOTH: TURKEY SENAT WHITE VOTE FILM SANDER STATE REPUBLICAN WOMEN ATTACK COURT ERDOGAN STUDENT CAMPAIGN AMERICAN ZIKA DEMOCRAT PROTEST NEWS ELECT RUSSIAN OBAMA VIDEO COUNTRI WEINER REFUGE

CONSERVATIVE ONLY: RUBIO ABEDIN TOLD BILL NEW LIVE GERMANI ISRAEL MEXICAN 2016 NFL GUN LAW RYAN MIGRANT TURKISH MEDIA FBI CARTEL ILLEG TWEET IOWA ISLAM PALESTINIAN CRIME MUSLIM BROADDRICK HILLARI 2017 CALL PERCENT DISCUSS PATRIOT ISRA CARSON BORDER GINGRICH SUPPORT CADDEL SCARBOROUGH KAIN JERUSALEM POLL CRUZ

2. 난민 문제

- Liberal과 Conservative에서 모두 발견됨
- Liberal - Russia, Tillerson, Assad
 - 러시아가 시리아 대통령 Assad를 지원하고 미국이 시리아 반군을 지원했음. 러시아가 지원해서 참전하려고 하자 미-러 평화협정을 맺기 위해 노력했던 사건을 뜻하고 있음
- Conservative - Migrant, Turkish
 - 터키 대통령 에르도안 대통령이 아사드 정부를 반대했고, 시리아 내전로 인해 터키의 난민이 급증했던 사건을 다루고 있음

LIBERAL ONLY: CHINA KOREA EU GOVERN BLACK EMISS EMOJI US ALBUM TAX MUSIC MARIJUANA TRANSGEND UPDAT CAR CARBON COAL PREGNANT YAHOO TRACK INTELLIG RUSSIA TILLERSON ASSAD TESLA HEALTH VACCIN MUSK ARCHIV INFECT SEASON DRIVER THRONE WATER READER POKÉMON FIRE PIPELIN TRIBE GUZMÁN MOVI SCIENC LOCHT SCHOOL CLIMAT SODA NUNE PRISON MOSQUITO

CONSERVATIVE ONLY: RUBIO ABEDIN TOLD BILL NEW LIVE GERMANI ISRAEL MEXICAN 2016 NFL GUN LAW RYAN MIGRANT TURKISH MEDIA FBI CARTEL ILLEG TWEET IOWA ISLAM PALESTINIAN CRIME MUSLIM BROADDRICK HILLARI 2017 CALL PERCENT DISCUSS PATRIOT ISRA CARSON BORDER GINGRICH SUPPORT CADDEL SCARBOROUGH KAIN JERUSALEM POLL CRUZ

3. 다른 정당 정치인 언급

- Liberal - Tillerson(보수주의)
→ Tillerson이 했던 Assad와 관련된 행동을 칭찬했음
- Conservative - Hillari(자유주의), Abedin(힐러리 보좌관)
→ 애버딘의 전 남편 앤서니 위너의 노트북에서 상당수의 힐러리와 애버딘 간의 이메일이 발견됐던 이메일 스캔들 자주 언급

LIBERAL ONLY: CHINA KOREA EU GOVERN BLACK EMISS EMOJI US ALBUM TAX MUSIC MARIJUANA TRANSGEND UPDAT CAR CARBON COAL PREGNANT YAHOO TRACK INTELLIG RUSSIA TILLERSON ASSAD TESLA HEALTH VACCIN MUSK ARCHIV INFECT SEASON DRIVER THRONE WATER READER POKÉMON FIRE PIPELIN TRIBE GUZMÁN MOVI SCIENC LOCHT SCHOOL CLIMAT SODA NUNE PRISON MOSQUITO

BOTH: TURKEY SENAT WHITE VOTE FILM SANDER STATE REPUBLICAN WOMEN ATTACK COURT ERDOGAN STUDENT CAMPAIGN AMERICAN ZIKA DEMOCRAT PROTEST NEWS ELECT RUSSIAN OBAMA VIDEO COUNTRI WEINER REFUGE

CONSERVATIVE ONLY: RUBIO ABEDIN TOLD BILL NEW LIVE GERMANI ISRAEL MEXICAN 2016 NFL GUN LAW RYAN MIGRANT TURKISH MEDIA FBI CARTEL ILLEG TWEET IOWA ISLAM PALESTINIAN CRIME MUSLIM BROADDRICK HILLARI 2017 CALL PERCENT DISCUSS PATRIOT ISRA CARSON BORDER GINGRICH SUPPORT CADDEL SCARBOROUGH KAIN JERUSALEM POLL CRUZ

4. 지카 바이러스

- Zika virus 오바마 케어 관련하여 좌우 대립
- Liberal - pregnant, health, vaccin, infect
- Conservative – ryan(공화당, 오바마케어 폐지 찬성 정치인)
→ zika virus는 당시 중요한 이슈

LIBERAL ONLY: CHINA KOREA EU GOVERN BLACK EMISS EMOJI US ALBUM TAX MUSIC MARIJUANA TRANSGEND UPDAT CAR CARBON COAL **PREGNANT** YAHOO TRACK INTELLIG **RUSSIA TILLERSON ASSAD TESLA HEALTH VACCIN** MUSK ARCHIV **INFECT** SEASON DRIVER THRONE WATER READER **POKÉMON** FIRE **PIPELIN TRIBE GUZMÁN** MOVI SCIENC LOCHT SCHOOL CLIMAT SODA NUNE PRISON **MOSQUITO**

BOTH: TURKEY SENAT **WHITE** VOTE FILM SANDER STATE REPUBLICAN **WOMEN** ATTACK COURT **ERDOGAN** STUDENT CAMPAIGN AMERICAN **ZIKA** DEMOCRAT PROTEST NEWS **ELECT** RUSSIAN **OBAMA** VIDEO COUNTRI **WEINER** REFUGE

CONSERVATIVE ONLY: RUBIO **ABEDIN** TOLD BILL NEW LIVE **GERMANI ISRAEL** MEXICAN 2016 NFL **GUN** LAW **RYAN** MIGRANT **TURKISH** MEDIA **FBI CARTEL ILLEG** TWEET IOWA **ISLAM** **PALESTINIAN CRIME MUSLIM** BROADDRICK **HILLARI** 2017 CALL PERCENT DISCUSS **PATRIOT** **ISRA CARSON BORDER GINGRICH** SUPPORT CADDEL SCARBOROUGH KAIN JERUSALEM POLL **CRUZ**

5. 주요 공약 차이

- 16~17년도(대선 시기) 기사에서 지지율 변화 및 후보와 당의 공약 관련 기사 많음

| 힐러리 클린턴과 도널드 트럼프의 정책 비교 | |
|--|--|
| 힐러리 클린턴 | 도널드 트럼프 |
| <ul style="list-style-type: none"> * 5년간 2750억 달러 인프라 투자(2차 세계대전 이후 최대 규모) * 미국 고용·임금상승으로 이어지지 않는 무역 협정에 반대 * TPP에도 동일 기준 적용 * 오바마케어 단점 보완 * 버핏세 도입 및 500만 달러 이상 소득자 세율구간 신설 * 석유·가스 대기업에 대한 세제혜택 철폐 * 동맹 강화/외환시장 개입에는 부정적 * 제재 강화/한국·일본과 방위협력 강화 * 불공정한 무역관행 대응 조치 * 센카쿠열도 도발에 반대 • IS 괴멸 위해 동맹국과의 광범위한 연합 주도 • 핵합의 지지 • 오바마 정부의 이민개혁 지지 • 총기 규제 강화/살상 능력 높은 무기·탄약 유통 금지 • 기후변화는 즉각적인 위협 • 10년내 에너지원 50% 청정에너지로 | <ul style="list-style-type: none"> • 성장률 3.5%/10년간 2500만 명 고용 창출 • 미국 우선 무역협상 • 국익 보호하지 못한 무역협정 거절 • TPP 반대, NAFTA 재협상 추진 • 오바마케어 폐지 • 법인세율 35% → 15%로 • 최고 소득세율 33%로 인하 • 주둔 미군 비용 전액 부담 • 중국이 북한 문제 해결 • 환율조작국 지정 • 동·남중국해에서 미군 전개 강화 • IS와 관련해 러시아와 협력 • 테러 관련국에서의 이민 금지 • 핵합의 파기할 것 • 멕시코 국경에 장벽 • 오바마 정부의 불법 이민자 사면 철회 • 개인이 무기 가질 권리 지킬 것 • 국가 안보 문제와 거리, 화석연료 개발 확대 • 파리 기후변화 협정 탈퇴 불사 |

5. 주요 공약 차이

- Liberal – black, transgender, climatee, carbonate, emission
→ 인권 및 환경 문제
- Conservative – migrant, mexican border, muslim
→ 안보 및 난민문제

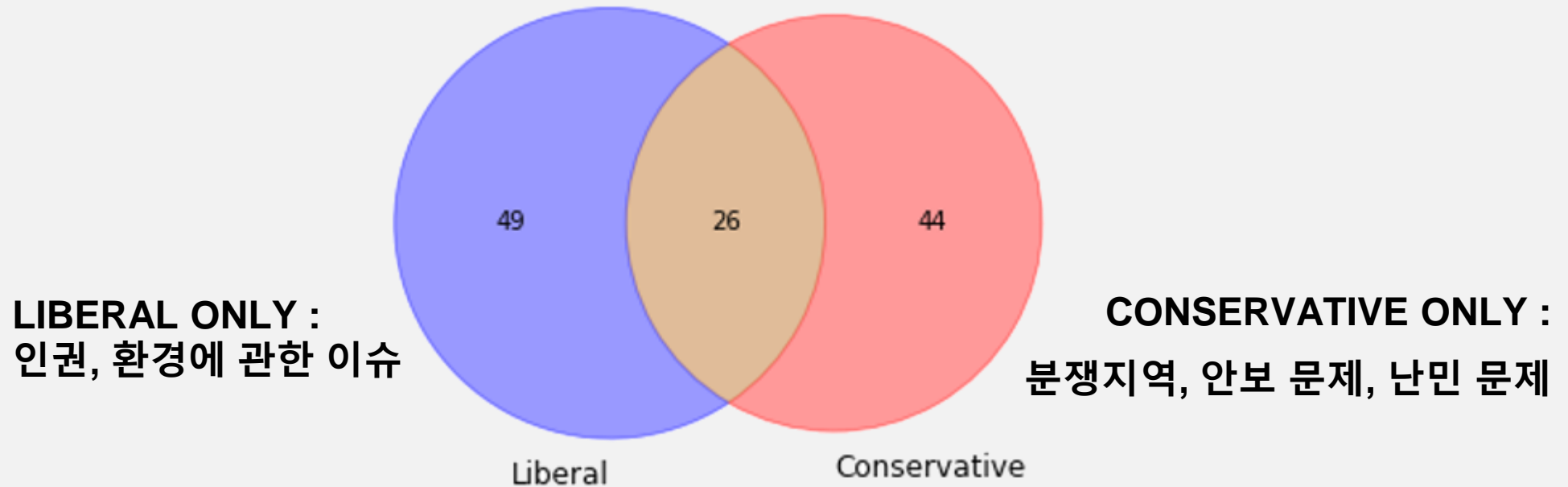
LIBERAL ONLY: CHINA KOREA EU GOVERN **BLACK EMISS** EMOJI US ALBUM TAX MUSIC MARIJUANA **TRANSGEND** UPDAT CAR **CARBON** COAL **PREGNANT** YAHOO TRACK INTELLIG **RUSSIA TILLERSON ASSAD** TESLA **HEALTH VACCIN** MUSK ARCHIV **INFECT** SEASON DRIVER THRONE WATER READER **POKÉMON** FIRE **PIPELIN** **TRIBE GUZMÁN** MOVI SCIENC LOCHT SCHOOL **CLIMAT** SODA NUNE PRISON MOSQUITO

BOTH: TURKEY SENAT **WHITE** VOTE FILM SANDER STATE REPUBLICAN **WOMEN** ATTACK COURT **ERDOGAN** STUDENT CAMPAIGN AMERICAN **ZIKA** DEMOCRAT PROTEST NEWS **ELECT** RUSSIAN **OBAMA** VIDEO COUNTRI **WEINER REFUGE**

CONSERVATIVE ONLY: RUBIO **ABEDIN** TOLD BILL NEW LIVE **GERMANI ISRAEL MEXICAN** 2016 NFL **GUN** LAW **RYAN** **MIGRANT** **TURKISH** MEDIA **FBI CARTEL ILLEG** TWEET IOWA **ISLAM PALESTINIAN CRIME** **MUSLIM** BROADDRICK **HILLARI** 2017 CALL PERCENT DISCUSS **PATRIOT ISRA** **CARSON** **BORDER** **GINGRICH** SUPPORT CADDEL SCARBOROUGH KAIN JERUSALEM POLL **CRUZ**

ANALYSIS

BOTH : 정치적 이슈(지카 바이러스, 오바마 케어 존폐 여부,
2016 경선 및 대선, 멕시코 국경 문제),
정치적 논란(힐러리 이메일 논란)



토픽 분석의 한계점

- ❖ **STEMMING**으로 인해 잘린 단어의 뜻을 파악하기 어려움.
- ❖ 반복적인 **Stopwords** 제거에도 남아있는 **Stopwords**
- ❖ **LDA** 결과로 단어에 주어진 **weigh**가 너무 작거나 같음(0.003이하)
- ❖ **미국에 대한 배경지식 부족**

CONCLUSION & FUTURE WORK

- 진보, 보수의 시사 방향성의 차이를 일부 확인할 수 있음
- 2016-17년도의 정치 흐름을 파악해, 진보, 보수의 입장 차이를 알아보기에는 기간이 광범위하다는 의견→ 특정 사건을 다루는 기사 데이터로 진행했다면 토픽이 더 명확했을 것
- Pre-processing의 중요성(stopwords)
- 중립적 언론사와 두 진영에서 쓰인 공통 단어와의 차이점을 비교할 수 있음

THANK YOU FOR LISTENING