



텍스트 마이닝을 이용한 금리 예측

Deciphering Monetary Policy Board Minutes through Text Mining
Approach: The Case of Korea

수진들과 아이들_박수진, 송은이, 윤호준, 김수진, 박창서

01 논문요약

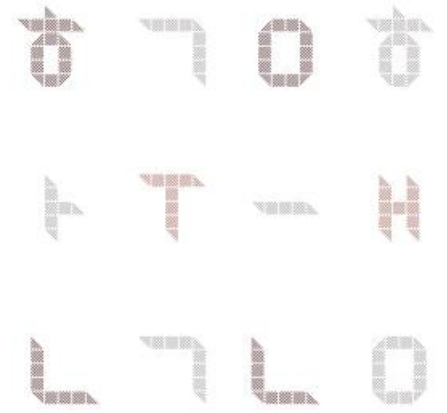
Ki Young Park 외2인은 금융통화위원회 의
사록에 나타난 비 정형적인 정보를 측정, 분석
하여 통화정책의 방향성을 예측하고자 하였다.

No. 2019-1

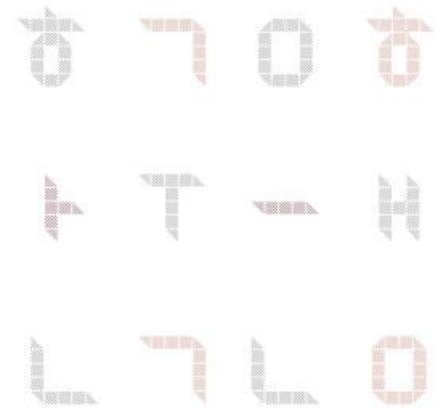
BOK Working Paper

Deciphering Monetary Policy Board
Minutes through Text Mining Approach:
The Case of Korea

Ki Young Park, Youngjoon Lee, Soohyon Kim



2019. 1



우리의 전략

크롤링 및 전처리

- 네이버기사(탑3 : 연합뉴스, 이데일리, 연합뉴스포맥스), 채권레포트, 한국은행 의사록 2011~2020 (10년치) (<-> 논문 2005~2017)
- 전처리 과정을 통해 n-gram 추출 및 라벨링

감성분석

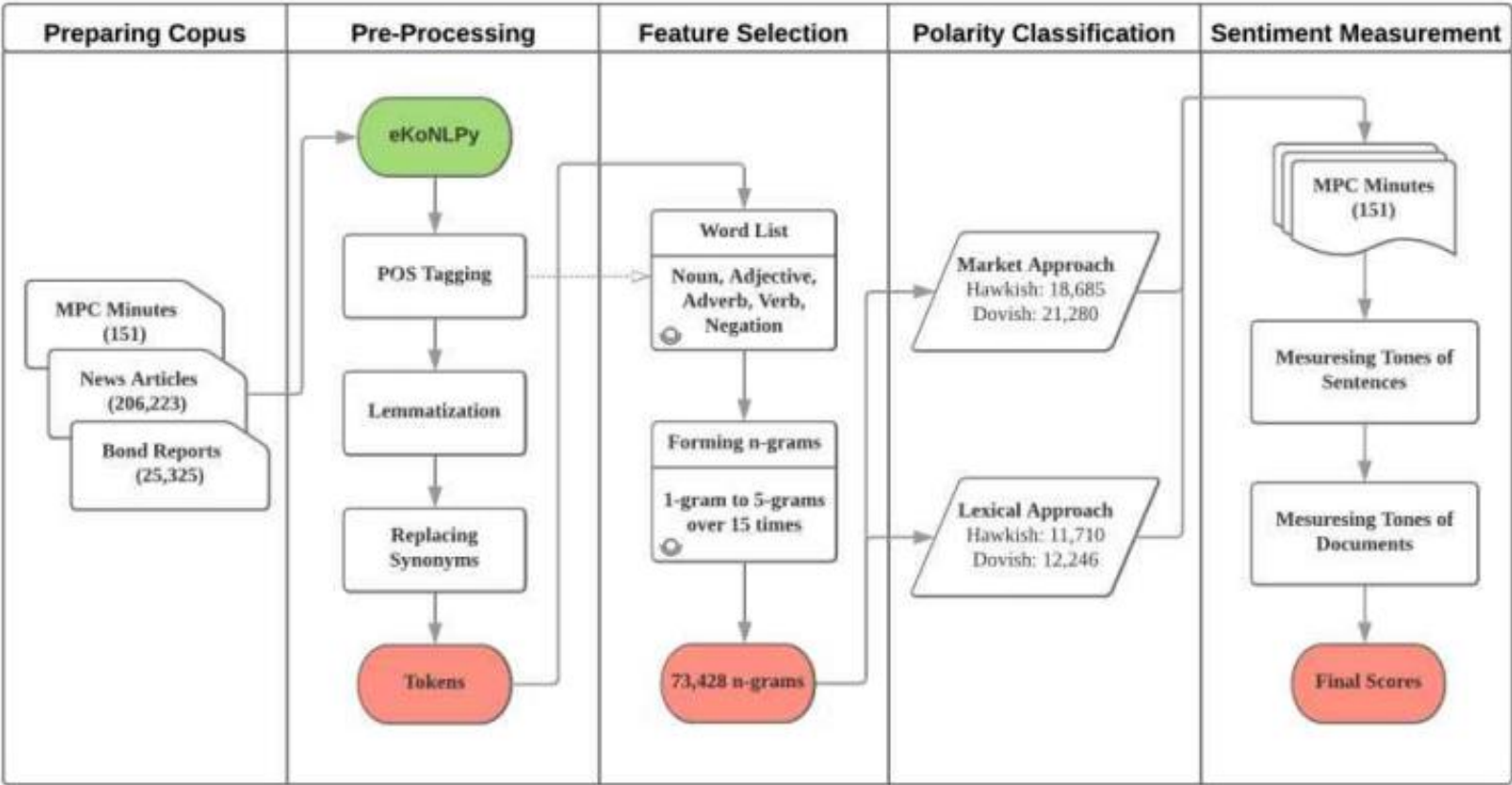
- n-gram의 극성 설정
명확한 분석을 위해
극성이 1.3 초과 : 긍정
극성이 0.76~1.3 : 중립
극성이 0.76 미만 : 부정
이를 이용해 문장과 의사록의 감성을 분석.
- 극성 분석에서 시장접근법을 활용.

결과값 비교

- 금통위 의사록 문장의 톤과 기준금리 간 상관관계 도출.
- 분석의 기간이 다를 경우에도 기준과 같은 결론이 나오는지 검증.
- 더 나아가, 현 코로나 사태에 분석 내용이 어떻게 달라지는지 확인.

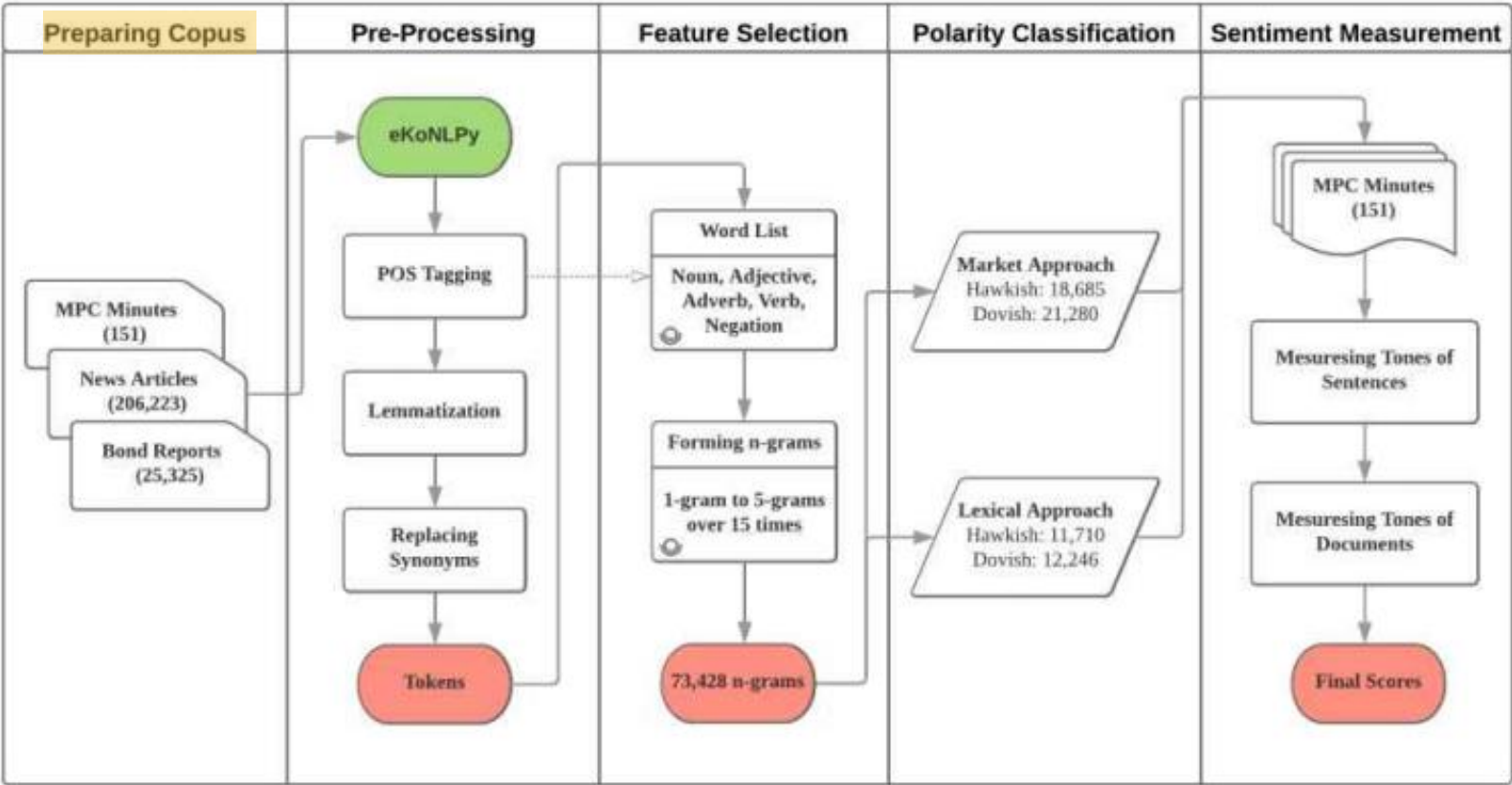
전체 과정

Figure 1. Procedure of Sentiment Analysis



Note: This figure, along with table 1, summarizes our discussion in section 3.

Figure 1. Procedure of Sentiment Analysis



Note: This figure, along with table 1, summarizes our discussion in section 3.

데이터 크롤링

For Monetary Policy Board Minutes

총 10년치
200개

To Download PDF files:

```
1 #금융위 의사록 다운로드 최종
2 import requests
3 from bs4 import BeautifulSoup
4 import datetime
5 from datetime import datetime
6 import re
7 for i in range(1,21): # 마지막 페이지 입력하기
8     URL = 'https://www.bok.or.kr/portal/bbs/B0000245/list.do?menuNo=200761&pageIndex='+str(i)
9     response = requests.get(URL)
10    soup = BeautifulSoup(response.text, "html.parser")
11    list_minutes = soup.select('main[id="content"] div[class="bdLine type2"] > ul > li')
12    for minutes in list_minutes:
13        title = minutes.select('a')[0].text.strip()
14        pdf_check = minutes.select('a')[2].text.strip()
15        pdf_check2 = minutes.select('a')[4].text.strip()
16        sub_link = minutes.select('a')[2].attrs['href']
17        sub_link2 = minutes.select('a')[4].attrs['href']
18        if re.findall(r'(?<=채#)', title)[0][-2] != '.': # (1111.11.11)형식이면
19            pdf_date = datetime.strptime(re.findall(r'(?<=채#)', title)[0][1:-1], "%Y.%m.%d")
20            pdf_date2 = pdf_date.strftime('%Y.%m.%d')
21        else:
22            pdf_date = datetime.strptime(re.findall(r'(?<=채#)', title)[0][1:-2], "%Y.%m.%d")
23            pdf_date2 = pdf_date.strftime('%Y.%m.%d')
24        if 'pdf' in pdf_check:
25            link = 'https://www.bok.or.kr' + sub_link
26        else:
27            link = 'https://www.bok.or.kr' + sub_link2
28        file_res = requests.get(link)
29        with open(f"{pdf_date2}.pdf", 'wb') as f:
30            f.write(file_res.content)
```

데이터 크롤링

To Extract Content from PDF Files:

총 10년치
200개

```
1 from tika import parser
2 from datetime import datetime, timedelta, date
3 f = open('mp_report.csv', 'w', newline='')
4 wr = csv.writer(f)
5 start_date = date(2011, 1, 1)
6 end_date = date(2020, 12, 31)
7 ds = start_date
8 while True:
9     try:
10         filename = f'./금융통화위원회 의사록/{ds.strftime("%Y.%m.%d")}.pdf'
11         parsed = parser.from_file(filename, requestOptions={'timeout': 120})
12         # print(ds)
13         # print(parsed['content'].strip())
14         wr.writerow([ds, parsed['content'].strip()])
15     except:
16         pass
17     ds += timedelta(1)
18     if ds == end_date:
19         break
20 f.close()
21 pd.read_csv('mp_report.csv', header=None)
```

데이터 크롤링

For Bond Analyst Reports

총 10년치
3,012개

```
1 # option 1
2 import requests
3 from bs4 import BeautifulSoup
4 import datetime
5 from datetime import datetime
6 !pip install openpyxl
7 import openpyxl
8 wb = openpyxl.Workbook()
9 sheet = wb.active
10 sheet.append(["날짜", "제목", "본문"])
11 for i in range(1,101): # 마지막 페이지 입력하기 101
12     URL = 'https://finance.naver.com/research/debenture_list.nhn?keyword=&brokerCode=&searchType=writeDate&writeFromDate=2011-01-01&writeToDate=2020-12-31'
13     response = requests.get(URL)
14     soup = BeautifulSoup(response.text, "html.parser")
15     list_minutes = soup.select('div[id="contentarea_left"] div[class="box_type_m"] tr')
16     for minutes in list_minutes:
17         try:
18             title=minutes.select('a')[0].text
19             sub_link = minutes.select('a')[1].attrs['href']
20             date = datetime.strptime('20'+ minutes.select('.date')[0].text, "%Y.%m.%d")
21             if str(sub_link).endswith('.pdf'):
22                 print(date.strftime('%Y.%m.%d'))
23                 print(title)
24                 print(sub_link)
25                 file_res = requests.get(sub_link)
26                 sheet.append([date.strftime('%Y.%m.%d'), title, sub_link])
27             else:
28                 pass
29         except:
30             pass
31 wb.save("bond_report.csv")
```


데이터 크롤링

총 10년치
3,012개

```
1 #링크만 pdf text 긁어오기
2 import os
3 from tika import parser
4 import pandas as pd
5 import csv
6 bond_report = pd.read_csv("bond_report.csv", header=0)
7 bond_report.columns = ['date', 'title', 'url']
8 content = []
9 for url in bond_report['url']:
10     try:
11         content.append(parser.from_file(url)['content'].replace('\n', ''))
12     except:
13         content.append("NA")
14     pass
15 bond_report['content'] = content
16 bond_report
17 bond_report.to_csv(r'/Users/soojinkim/kdigital/mini_project_1/bond_report_df.csv')
```

일드 픽업 이제는 장기물이 답이다

미래에셋대우



20.12.18

430

10년 발행 감소, 초장기 공급부담 완화

하나금융투자



20.12.18

208

데이터 크롤링

N | 금리



SBS Biz PICK | 10시간 전 | 네이버뉴스

올 1학기 학자금 대출금리 1.7%...지난 학기보다 0.15%p 인하

올해 1학기 대학 학자금 대출금리가 지난 학기인 작년 2학기보다 인하됩니다. 교육부와 한국장학재단은 올해 1학기 학자금 대출금리를 지난 학기 연 1.85%에서 0.1...



학자금 대출금리 내려간다...작년보다... 서울경제 PICK | 9시간 전 | 네이버뉴스

학자금 대출금리, 작년보다 0.15%p 내린 연 1... 한겨레 | 8시간 전 | 네이버뉴스

1학기 대학 학자금 대출금리 1.7%...작년 2... 연합뉴스 | 10시간 전 | 네이버뉴스

1학기 대학 학자금 대출금리 1.7%...작년 2학기... SBS | 10시간 전 | 네이버뉴스

관련뉴스 21건 전체보기 >

디지털타임스 | 3시간 전 | 네이버뉴스

법정최고금리 인하에 제2금융권 근심

금융당국 입김으로 여전업체 기존 대출도 인하 가능성 커 법정최고금리 인하로 저신용자 금융소외 현상 더 커질 우려 올해 하반기 법정최고금리 인하 예정으로, 카...



새해 바뀌는 저축은행 제도...법정 최고금리↓ 굿모닝경제 | 12시간 전

총 10년치
223,562개

연합 인포맥스, 다른 언론사 다르게

데이터 크롤링

For Naver News Articles

```
1 import openpyxl
2 wb = openpyxl.Workbook()
3 sheet = wb.active
4 sheet.append(["날짜", "기사명", "본문"])
5 #은이날 수정
6 from bs4 import BeautifulSoup
7 import requests
8 import csv
9 import re
10 import datetime
11 start_date = datetime.datetime(2013,1,1)
12 end_date = datetime.datetime(2014,1,1)
13 total_days = int((end_date-start_date).days)
14 ds=start_date #시작일자
15 errorcount=0
16 for day in range(1,total_days+1):
17     page = 1
18     while True:
19         headers = {'user-agent':'Mozilla/5.0','cookie':'news_office_checked=1001,1018,2227'}
20         url = f"https://search.naver.com/search.naver?&where=news&query=%EA%B8%88%EB%A6%AC&sm=tab_pge&sort=0&photo=0&field=0&report"
21         response = requests.get(url,headers=headers)
22         soup = BeautifulSoup(response.text,"html.parser")
23         li_list = soup.select(".list_news > li")
```

총 10년치
223,562개

연합 인포맥스, 다른 언론사 다르게

데이터 크롤링

```
24     for li in li_list:
25         try:
26             # print(ds.strftime('%Y.%m.%d'))
27             if li.select('div .news_area a[href]')[0]['href'] == 'http://www.einfomax.com':
28                 title = li.select('div .news_area a[title]')[0].text
29                 article_url = li.select('div .news_area a[title]')[0]['href']
30                 article_response = requests.get(article_url, headers=headers)
31                 article_soup = BeautifulSoup(article_response.text, "html.parser")
32                 article = article_soup.find(id = 'article-view-content-div')
33                 body = re.sub(r'[\n#t]', '', article.text.strip())
34             else:
35                 title = li.select('div .news_area a[title]')[0].text
36                 article_url = li.select('div .news_area a[href]')[1]['href']
37                 article_response = requests.get(article_url, headers=headers)
38                 article_soup = BeautifulSoup(article_response.text, "html.parser")
39                 article = article_soup.find(id = 'articleBodyContents')
40                 body = article.text.strip()
41             # print(title)
42             # print(body)
43             # print("")
44             sheet.append([ds.strftime('%Y.%m.%d'), title, body])
45         except:
46             errorcount+=1
47             button=soup.select('.sc_page > a')[1]['aria-disabled']
48             page+=10
49             if button == 'true':
50                 ds+=datetime.timedelta(1)
51                 break
52     wb.save("2020.csv")
```

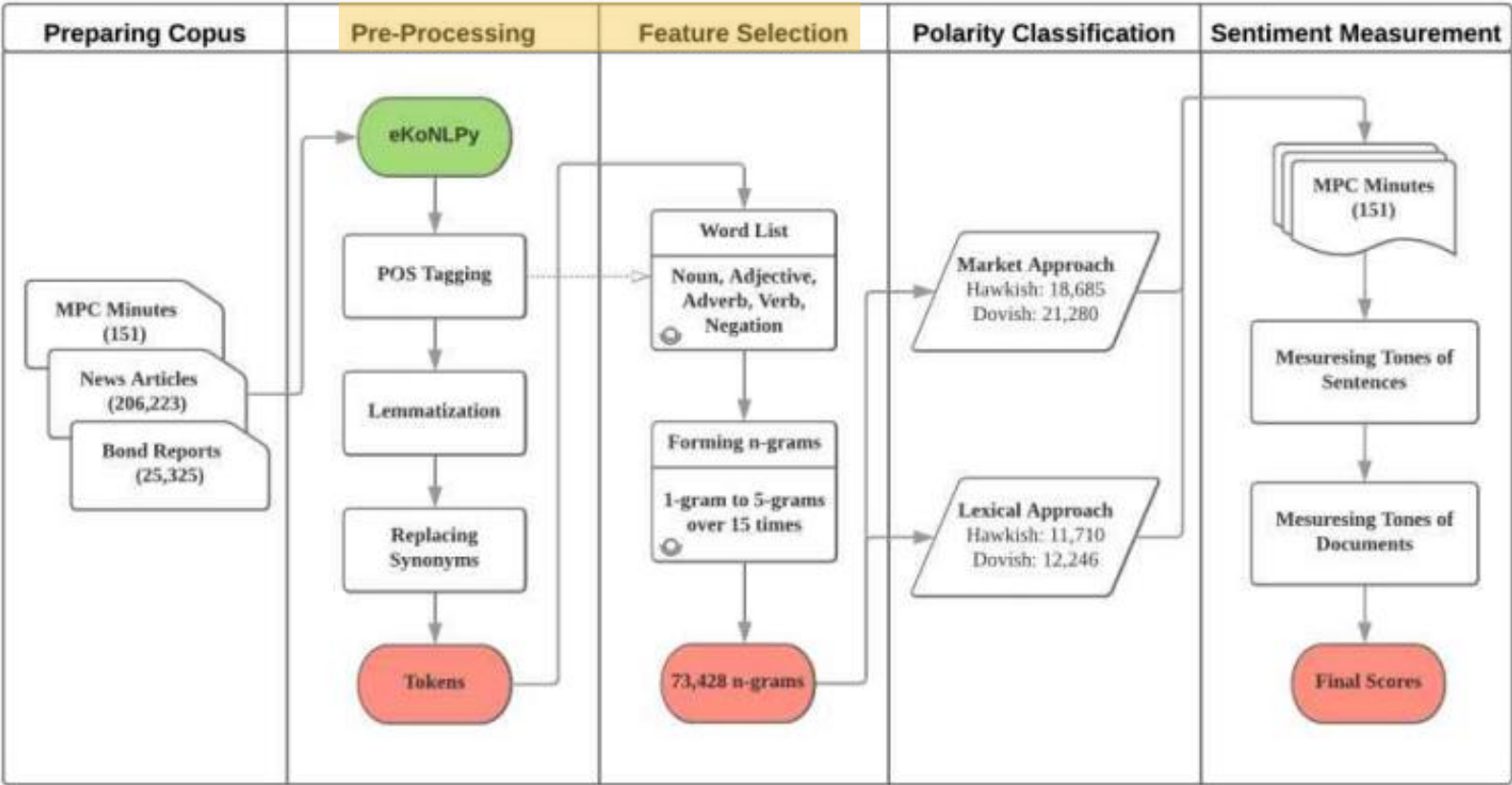
총 10년치
223,562개

연합 인포맥스, 다른 언론사 다르게

1년 X 10개 취합

전처리 과정

Figure 1. Procedure of Sentiment Analysis



Note: This figure, along with table 1, summarizes our discussion in section 3.

전처리 과정

eKo(nomic)NLPy

eKoNLPy is a Korean NLP Python Library for Economic Analysis, which supports Korean Language.

KoNLPy 의 Mecab tagger를 기반으로 경제관련 전문용어, 금융기관, 기업명 등을 하나의 명사로 분류하도록 후처리 기능을 추가.

통화정책(Monetary Policy)의 어조(Hawkish/Dovish)를 판단할 수 있는 Sentiment Analysis 기능 포함.

경제의 불확실성(Uncertain/Stable)을 판단할 수 있는 Economic Uncertainty Analysis 기능 포함.

경제 문서의 주제를 분류할 수 있는 Topic Analysis 기능 포함.

한국은행/NNG, 이/JKS, 12/SN, 일/NNG, 금융통화위원회/NNG,
금통위/NNG, 회의/NNG, 를/JKO, 열/VV, 고/EC, 기준금리/NNG,
를/JKO, 현행/NNG, 연/NNG, 1/SN, ./SW, 50/SN, %/SW, 로/JKB,
동결/NNG, 했/XSV, 다/EC (Lee, Kim and Park, 2019a)

전처리 과정

```
1
2 tokens = mpck.tokenize(mp_report['content'])
3 ngrams = mpck.ngramize(tokens)
4 print(ngrams)
5 #0/개 a=1
```

['소비자/NNG:물가/NNG:상승률/NNG:물가/NNG:안정/NNG', '국제/NNG:원자재/NNG:가격/NNG:변동성/NNG:확대/NNG', '물가/NNG:상승률/NNG:농산물/NNG:가격/NNG:상승/NNG', '글로벌/NNG:경기/NNG:회복/NNG:수요/NNG:증가/NNG', '소비자/NNG:물가/NNG:국제/NNG:유가/NNG:상승/NNG', '세계/NNG:경제/NNG:신흥국/NNG:경제/NNG:호조/NNG', '선행/NNG:지수/NNG:동행/NNG:지수/NNG:하락/NNG', '가격/NNG:오르/VV:물가/NNG:상승/NNG:압력/NNG', '국제/NNG:금융시장/NNG:불확실성/NNG:확대/NNG', '중소기업/NNG:대출/NNG:연체율/NNG:상승/NNG', '주택담보대출/NNG:주택/NNG:거래/NNG:늘/VV', '통화정책/NNG:경제/NNG:건조/NNG:성장/NNG', '전망/NNG:선진국/NNG:경기/NNG:회복/NNG', '선진국/NNG:경기/NNG:회복/NNG:강화/NNG', '원리금/NNG:상환/NNG:부담/NNG:증가/NNG', '소비자/NNG:물가/NNG:수요/NNG:압력/NNG', '상승률/NNG:전망/NNG:물가/NNG:안정/NNG', '소비자/NNG:물가/NNG:경기/NNG:상승/NNG', '국제/NNG:원자재/NNG:가격/NNG:상승/NNG', '재정/NNG:문제/NNG:지정학/NNG:위험/NNG', '자산/NNG:가격/NNG:변동성/NNG:확대/NNG', '자본/NNG:유출입/NNG:규제/NNG:강화/NNG', '경기/NNG:회복/NNG:경기/NNG:회복/NNG', '가계/NNG:부채/NNG:문제/NNG:심각/NNG', '실물/NNG:자산/NNG:가격/NNG:하락/NNG', '기대/NNG:기업/NNG:실적/NNG:개선/NNG', '가격/NNG:상승/NNG:물가/NNG:불안/NNG', '수요/NNG:물가/NNG:상승/NNG:압력/NNG', '가계/NNG:대출/NNG:증가/NNG:확대/NNG', '지표/NNG:물가/NNG:상승/NNG:압력/NNG', '물가/NNG:상승/NNG:압력/NNG:증대/NNG', '물가/NNG:상승/NNG:압력/NNG:해소/NNG', '수출/NNG:세계/NNG:경제/NNG:회복/NNG', '물가/NNG:안정/NNG:노력/NNG:강화/NNG', '실물/NNG:경제/NNG:활동/NNG:위축/NNG', '경기/NNG:선행/NNG:지수/NNG:하락/NNG', '경기/NNG:회복/NNG:기대/NNG:증가/NNG', '금리/NNG:상승/NNG:영향/NNG:상승/NNG', '은행/NNG:시장/NNG:금리/NNG:상승/NNG', '물가/NNG:상승률/NNG:전망/NNG:높/VV', '민간/NNG:소비/NNG:증가율/NNG:낮/VV', '물가/NNG:상승/NNG:압력/NNG:크/VV', '높/VV:물가/NNG:상승/NNG:압력/NNG', '인플레이션/NNG:기대/NNG:상승/NNG', '기대/NNG:인플레이션/NNG:안정/NNG', '농축수산물/NNG:가격/NNG:급등/NNG', '인플레이션/NNG:기대/NNG:증폭/NNG', '기대/NNG:인플레이션/NNG:상승/NNG', '농축수산물/NNG:가격/NNG:상승/NNG', '유로존/NNG:재정위기/NNG:확산/NNG', '인플레이션/NNG:기대/NNG:확산/NNG', '인플레이션/NNG:기대/NNG:확대/NNG', '경제/NNG:인플레이션/NNG:위험/NNG', '인플레이션/NNG:위험/NNG:경계/NNG', '부동산/NNG:전세가격/NNG:상승/NNG', '비농업/NNG:고용자수/NNG:증가/NNG', '공공측면/NNG:물가/NNG:충격/NNG', '거시정책/NNG:물가/NNG:안정/NNG', '취약/NNG:재무구조/NNG:개선/NNG', '경착륙/NNG:가능성/NNG:우려/NNG', '선진국/NNG:신흥국/NNG:성장/NNG', '주택/NNG:전세가격/NNG:상승/NNG', '투자자금/NNG:유입/NNG:확대/NNG', '전망/NNG:통화정책/NNG:완화/NNG', '국제/NNG:금융시장/NNG:불안/NNG', '안전통화/NNG:선호/NNG:약화/NNG', '재정거래/NNG:유인/NNG:축소/NNG', '신흥국/NNG:경기/NNG:회복/NNG', '인플레이션/NNG:기대/NNG:높/VV', '소비자/NNG:물가/NNG:하락/NNG', '거품/NNG:부동산/NNG:거품/NNG', '금리/NNG:상승/NNG:부작용/NNG', '민간/NNG:취업자/NNG:증가/NNG', '풍부/NNG:유동성/NNG:상승/NNG', '회복/NNG:모멘텀/NNG:강화/NNG', '재무/NNG:건전성/NNG:강화/NNG', '소비자/NNG:물가/NNG:상승/NNG', '유동성/NNG:관리/NNG:강화/NNG', '소비자/NNG:물가/NNG:급등/NNG', '지정학/NNG:위험/NNG:완화/NNG', '식료품/NNG:가격/NNG:인상/NNG', '과잉/NNG:유동성/NNG:축소/NNG', '식료품/NNG:가격/NNG:상승/NNG',

전처리 과정

실제 의사록 전처리

```
1 # 2~3번 섹션만 가져오기 - 2
2 import re
3
4 def preprocess_minutes(minutes):
5
6     pos = re.search('(.(?국내외\s?경제\s?동향.?과 관련하여,?\|가#\s?+경제전망.*\|가#\s?국내외 경제동향 및 평가)\n?\s*일부 위원은', minutes)
7     s1 = pos.start() if pos else -1
8     pos = re.search('(.(?외환.?국제금융\s?동향.?과 관련하여,?\|나#\s?외환.국제금융\s?(및 금융시장)?\s?동향)\n?\s*(일부 위원은|대부분의')
9     s2 = pos.start() if pos else -1
10    pos = re.search('(.(?금융시장\s?동향.?과 관련하여,?\|다#\s?금융시장\s?동향)\n?\s*일부 위원은', minutes, re.MULTILINE)
11    s3 = pos.start() if pos else -1
12    pos = re.search('(((다\|라#\s?))\s?)?.?통화정책\s?방향.?에 관한 토론,?\|이상과 같은 의견\s?교환을 바탕으로.*통화정책\s?방향.*에.*토론.*')
13    s4 = pos.start() if pos else -1
14    pos = re.search('(4#\s?정부측 열석자 발언.*)\n?', minutes, re.MULTILINE)
15    s5 = pos.start() if pos else -1
16    pos = re.search('(.\s*) 한국은행 기준금리 결정에 관한 위원별 의견\s?개진\|이상과 같은 토론에 이어 .* 관한 위원별 의견개진이 있었음')
17    s6 = pos.start() if pos else -1
18    positer = re.finditer('(4#\s?.\s*\s?)\s?((심의결과|토의결론))\n?', minutes, re.MULTILINE)
19    s7 = [pos.start() for pos in positer if pos.start() > s6]
20    s7 = s7[0] if s7 else -1
21
```


전처리 과정

For Monetary Policy Board Minutes

```
import csv
import pandas as pd
from ekonlpy.sentiment import MPCK
mpck = MPCK()

mp_report_ngram = pd.read_csv('mp_report_ngram.csv', header=0)
ngram_all=[]

for i in range(200):
    text = mp_report_ngram['content'][i]
    tokens = mpck.tokenize(text)
    ngrams = mpck.ngramize(tokens)

    if (tokens == []) == True:
        ngrams_all.append("")
    else:
        ngrams_all.append(",".join(tokens) + ",".join(ngrams))

for j in range(200):
    mp_report_ngram['ngram'][j] = ngram_all[j]

mp_report_ngram.to_csv(r'/Users/yunhojun/kdigital/mp_report_ngram1.csv')
```

전처리 과정

| | date | title | content | ngram |
|-----|------------|-------|--|---|
| 0 | 2011.01.13 | NaN | 일부 위원은 최근 외국인 채권투자자·금의 순유출로 외환수급사정이 악화되었음에도 ... | 외국인/NNG, 채권/NNG, 금/NNG, 순유출/NNG, 외환/NNG, 수급/NN... |
| 1 | 2011.02.11 | NaN | 일부 위원은 향후 미국의 통화정책 기조가 바뀔 경우 해외자금 유출입 흐름이 급격... | 통화정책/NNG, 기조/NNG, 바뀔/VV, 해외/NNG, 자금/NNG, 유출입/N... |
| 2 | 2011.03.10 | NaN | 일부 위원은 최근 유로화가 달러화에 대 해 상대적으로 강세를 보이고 있는 것이 달... | 유로/NNG, 달러/NNG, 대/NNG, 상대적/VAX, 강세/NNG, 보/VV, ... |
| 3 | 2011.04.12 | NaN | 일부 위원은 최근 역외 시장에서 대NDF 규모 선물환 매도로 인해 단기외채가 급증... | 역외시장/NNG, 대/NNG, 규모/NNG, 선물환/NNG, 매도/NNG, 인하/V... |
| 4 | 2011.05.13 | NaN | 일부 위원은 최근 외환보유액 적정수준에 대한 논란과 관련하여 에서 적정수준 산정시... | 외환보유액/NNG, 적정/NNG, 수준/NNG, 논란/NNG, 에서/NNG, 적정/... |
| ... | ... | ... | ... | ... |
| 96 | 2020.05.28 | NaN | 일부 위원은 코로나19 사태 이전 풍부한 유동성에 힘입어 미 증시가 사상 최 고치... | 코로나/NNG, 사태/NNG, 이전/NNG, 풍부/NNG, 유동성/NNG, 힘입/V... |
| 97 | 2020.07.16 | NaN | 일부 위원은 최근 외국인의 국내채권 투자를 투자자 형태별로 살펴보면 공공 부문이 ... | 외국인/NNG, 채권투자/NNG, 형태/NNG, 살펴보/VV, 공공/NNG, 단기/... |
| 98 | 2020.08.27 | NaN | 일부 위원은 비록 일부 위험요인이 남아있지만 국내외 금융시장과 외화자금 시장이 대... | 비록/MAG, 위험요인/NNG, 남/VV, 국내외/NNG, 금융시장/NNG, 외화자... |
| 99 | 2020.10.14 | NaN | 일부 위원은 코로나19 사태에도 불구하고 중국 금융시장으로 외국인 증권투 자자금이... | 코로나/NNG, 사태/NNG, 불구/NNG, 금융시장/NNG, 외국인/NNG, 증권... |
| 100 | 2020.11.26 | NaN | 일부 위원은 글로벌 투자은행들이 내년에도 달러화의 약세 추세가 대체로 이 어질 것... | 글로벌/NNG, 투자은행/NNG, 달러/NNG, 약세/NNG, 추세/NNG, 대체로... |

101 rows × 4 columns

전처리 과정

For Bond Analyst Reports

```
1 import pandas as pd
2 import csv
3 from ekonlpy.tag import Mecab
4 from ekonlpy.sentiment import MPCK
5 mpck = MPCK()
6 bond_report=pd.read_csv("bond_report_final.csv", header=0, encoding='utf-8')
7 bond_report.insert(4, 'tokens_ngrams', None)
8 bond_report
9 for row in range(len(bond_report)) : #레포트 개수
10     try:
11         text = list([sentence for sentence in str(bond_report.loc[row, "content"]).split(".")])
12         new_content=[]
13         for sentence in text: # 한 레포트 안의 문장 개수
14             tokens = mpck.tokenize(sentence)
15             ngrams = mpck.ngramize(tokens)
16             new_content+=tokens
17             new_content+=ngrams
18         bond_report['tokens_ngrams'][row]=new_content
19     except:
20         bond_report['tokens_ngrams'][row]="None"
21
22 bond_report.to_csv('bond_report_ngramized.csv', encoding='utf-8', index=False)
```

전처리 과정

| 1 | bond_report_new | | | |
|------|-----------------|---------------------------------|---|---|
| | date | title | content | ngrams |
| 0 | 2020.12.30 | 원점회귀原點回歸 | 2013년 0월 0일 원점회귀(原點回歸) Issue Comment ... | ['원점/NNG', '회귀/NNG', '原點/NNG', '回歸/NNG', '의현/N... |
| 1 | 2020.12.23 | 포트폴리오 다변화로 가는 여정 | Microsoft Word - 2020122217163761.docx 8 글로벌... | ['글로벌/NNG', '채권/NNG', '채권투자/NNG', '전략/NNG', '포... |
| 2 | 2020.12.23 | 2021년 전망 업데이트 | 2013년 0월 0일 2021년 전망 업데이트 Issue Comme... | ['전망/NNG', '의현/NNG', '보/VV', '조사/NNG', '자료/NNG... |
| 3 | 2020.12.18 | 일드 픽업 이제는 장거물이 답이다 | Microsoft Word - 2020121719393409.docx 그림 1... | ['pf/NNG', '벤치마크/NNG', '수익률/NNG', '추이/NNG', '주... |
| 4 | 2020.12.18 | 10년 발행 감소, 초장기 공급부담 완화 | 슬라이드 1Compliance Notice 본 조사자료는 고객의 투자에 정보를 제공... | ['슬라이드/NNG', '보/VV', '조사/NNG', '자료/NNG', '고객/N... |
| ... | ... | ... | ... | ... |
| 3007 | 2011.01.05 | 국고채 물량부담 지속 및 통화정책 경계감으로 인해 .. | Microsoft Word - 110105_DailyTrading_OCS 본 조사... | ['보/VV', '조사/NNG', '분석/NNG', '자료/NNG', '신회/NNG... |
| 3008 | 2011.01.04 | 미국 채권시장 약세의 영향으로 인해 약세가 진행될 .. | Microsoft Word - 110104_DailyTrading_OCS 본 조사... | ['보/VV', '조사/NNG', '분석/NNG', '자료/NNG', '신회/NNG... |
| 3009 | 2011.01.03 | 추가적인 금리상승이 불가피한 상황 | Microsoft Word - HMC_Bond weekly_110103.docHMC... | ['추가/NNG', '금리/NNG', '상승/NNG', '불가피/NNG', '채권/... |
| 3010 | 2011.01.03 | Level-up 을 위한 Warming | Microsoft Word - 110103 본드브리프_최종_.doc ... | ['본드/NNG', '브리프/NNG', '최종/NNG', '레벨업/NNG', '위하... |
| 3011 | 2011.01.03 | 국고채 3년 입찰물량 부담으로 인해 약세가 진행될 것.. | Microsoft Word - 110103_DailyTrading_OCS 본 조사... | ['보/VV', '조사/NNG', '분석/NNG', '자료/NNG', '신회/NNG... |

3012 rows × 4 columns

전처리 과정

For News Articles

```
1 import pandas as pd
2 import re
3 import csv
4 from konlpy.tag import Mecab
5 from konlpy.sentiment import MPCK
6 mpck = MPCK()
7
8 #2011~2020까지 10회 반복(파일 당 1년)
9
10 article=pd.read_csv("new_article_2016.csv", header=0, encoding='utf-8')
11 article.insert(3,'tokens_ngrams',None)
12 for i in range(len(article)):
13     print(i)
14     try:
15         piece = str(article.loc[i, "본문"])
16         piece = tidy_sentences(piece)
17         new_content=[]
18         for sentence in piece[0]:
19             if ("// flash 오류를 우회하기 위한 함수 추가 function _flash_removeCallback" in sentence) == True:
20                 sentence = re.sub("// flash 오류를 우회하기 위한 함수 추가 function _flash_removeCallback","",sentence)
21             else:
22                 pass
23             tokens = mpck.tokenize(sentence)
24             ngrams = mpck.ngramize(tokens)
25             new_content+=tokens
26             new_content+=ngrams
27         article['tokens_ngrams'][i]=new_content
28     except:
29         article['tokens_ngrams'][i]="None"
30
31 article.to_csv('naver_ngram_2016.csv', encoding='utf-8', index=False)
```

전처리 과정

| 날짜 | | 기사명 | 본문 | tokens_ngrams |
|--------|------------|--|--|---|
| 0 | 2011.01.01 | <표> 은행 정기예금 금리(31일 기준) | // flash 오류를 우회하기 위한 함수 추가\function_flash_re... | NaN |
| 1 | 2011.01.01 | [2011 금융]㉠가계부채 '체질 바꾸기' 가능할까 | // flash 오류를 우회하기 위한 함수 추가\function_flash_re... | [가계/NNG', '부채/NNG', '체질/NNG', '개 선/NNG', '대책/N... |
| 2 | 2011.01.01 | 국내 주식펀드 수익률 7주째 '플러스' | // flash 오류를 우회하기 위한 함수 추가\function_flash_re... | ['해외/NNG', '주식펀드/NNG', '마이너 스/NNG', '전환/NNG', '... |
| 3 | 2011.01.01 | [2011 뉴이머징]㉡가속페달에서 발 뻗 중국 | // flash 오류를 우회하기 위한 함수 추가\function_flash_re... | ['과열/NNG', '우려/NNG', '안정/NNG', '성 장/NNG', '괴하/V... |
| 4 | 2011.01.01 | [2011경제]㉢연구소보다 증권사 전망이 더 좋네 | // flash 오류를 우회하기 위한 함수 추가\function_flash_re... | ['최고/NNG', '새해/NNG', '경제/NNG', '물/VV', '경제/NNG... |
| ... | ... | ... | ... | ... |
| 220777 | 2020.12.31 | 연말 원화 자금시장 안정의 '숨은 공'...韓紙 5조원 돌었다 | (서울=연합인포맥스) 노현우 기자 = 단기 자금시장이 연말에 도 안정세를 보이는 가운데... | ['기자/NNG', '단기자금시장/NNG', '안 정/NNG', '보/VV', '가운... |
| 220778 | 2020.12.31 | "올해 세계 상장기업 시가총액 1경6천300조원 증 가...중앙은행 완화 덕" | (서울=연합인포맥스) 문정현 기자 = 올해 세계 증시가 상승 하면서 상장 기업의 시가... | ['기자/NNG', '세계/NNG', '증시/NNG', '상 승/NNG', '하/XS... |
| 220779 | 2020.12.31 | 비앙코 "2021년, 30년만에 물가 돌아오는 해 될 것" | (뉴욕=연합인포맥스) 우성문 통신원 = 비앙코 리서치의 짐 비앙코 대표가 2021년... | ['뉴욕/NNG', '짐/NNG', '물가/NNG', '다 시/MAG', '돌아오/V... |
| 220780 | 2020.12.31 | <딜링룸 백브리핑> NFL 선수 오쿵, 비트코인으로 연봉 수령 | (서울=연합인포맥스) 국제경제부 = 미국 프로축구(NFL) 캐 롤라이나 팬서스의 공... | ['국제/NNG', '경제부/NNG', '프로/NNG', '축구/NNG', '스팀/... |
| 220781 | 2020.12.31 | <뉴욕은 지금> 브렉시트와 디스플레이션의 수출 | (뉴욕=연합인포맥스) 2020.12.24 vs 2020.12.30 의미 없는 숫자의... | ['뉴욕/NNG', '의미/NNG', '없/VV', '숫 자/NNG', '나열/NNG... |

220782 rows × 4 columns

전처리 과정

Merge 전 각 DataFrame 정리:

date, title, content, ngrams 컬럼, 날짜 형식 통일

금통위 의사록

```
1 #mp_report_ngram.csv 데이터프레임 정리:
2 mp=pd.read_csv("mp_report_ngram.csv", header=0, encoding='utf-8')
3 #컬럼명 바꾸기
4 mp = mp.rename(columns={'ngram': 'ngrams'})
5 #추가된 Unnamed index 컬럼 제거
6 mp.drop(mp.filter(regex="Unnamed: 0"),axis=1, inplace=True)
7 mp.drop(mp.filter(regex="Unnamed: 0.1"),axis=1, inplace=True)
8 # 금통위의사록 데이터프레임은 title 컬럼 없으므로 추가해줄
9 mp.insert(1,'title',None)
10 # 날짜 형식 yyyy.mm.dd로 변환
11 for i in range(len(mp)):
12     mp['date'][i]=re.sub("-", ".",mp['date'][i])
13
14 mp
```

금통위 의사록, 채권 레포트, 뉴스 기사 모두 동일한 과정을 거친다.

```
1 #네이버 뉴스 데이터프레임과 채권리포트 데이터프레임, mp 데이터프레임 merge
2 ngram_join = pd.concat([Merged_na, bond_report, mp], ignore_index=True)
3 ngram_join
```

전처리 과정

| | date | updown | title | content | ngrams |
|--------|---------------------|--------|--|--|---|
| 0 | 2011-01-03 00:00:00 | DOWN | [마켓in]새해 외화채발행 위축될듯.."만기↓·조달금리↑" | // flash 오류를 우회하기 위한 함수 추가\function _flash_re... | ['외표채/NNG', '만기도래액/NNG', '추정/NNG', '사기업/NNG', ... |
| 1 | 2011-01-03 00:00:00 | DOWN | [마켓in][채권폴]"금리 되돌림..캐리수요 유입 관심" | // flash 오류를 우회하기 위한 함수 추가\function _flash_re... | ['국고채/NNG', '발행/NNG', '불확실성/NNG', '해소/NNG', '되... |
| 2 | 2011-01-03 00:00:00 | DOWN | [마켓in]동부증권 "3년물 이 금리상승 견인" | // flash 오류를 우회하기 위한 함수 추가\function _flash_re... | ['시장/NNG', '기사/NNG', '프리미엄/NNG', '시장/NNG', '줄고... |
| 3 | 2011-01-03 00:00:00 | DOWN | "5% 고성장" 방점..한은 기준금리 정상화 '헛로' | // flash 오류를 우회하기 위한 함수 추가\function _flash_re... | ['성장/NNG', '없이/MAG', '일자리/NNG', '복지/NNG', '재정/... |
| 4 | 2011-01-03 00:00:00 | DOWN | [마켓in]하나대투증권 "올해 채권금리 상승..국고3년 4%대" | // flash 오류를 우회하기 위한 함수 추가\function _flash_re... | ['시장/NNG', '기사/NNG', '프리미엄/NNG', '시장/NNG', '줄고... |
| ... | ... | ... | ... | ... | ... |
| 202046 | 2020-12-31 00:00:00 | UP | 연말 원화 자금시장 안정의 '숨은 공'...韓銀 5조원 풀었다 | (서울=연합인포맥스) 노현우 기자 = 단기 자금시장이 연말에도 안정세를 보이는 가운데... | ['기자/NNG', '단기자금시장/NNG', '안정/NNG', '보/VV', '가운... |
| 202047 | 2020-12-31 00:00:00 | UP | "올해 세계 상장기업 시가총액 1경6천300조원 증가...중앙은행 완화 덕" | (서울=연합인포맥스) 문정현 기자 = 올해 세계 증시가 상승하면서 상장 기업의 시가... | ['기자/NNG', '세계/NNG', '증시/NNG', '상승/NNG', '하/XS... |
| 202048 | 2020-12-31 00:00:00 | UP | 비앙코 "2021년, 30년만에 물가 돌아오는 해 될 것" | (뉴욕=연합인포맥스) 우성문 통신원 = 비앙코 리서치의 짐 비앙코 대표가 2021년... | ['뉴욕/NNG', '짐/NNG', '물가/NNG', '다시/MAG', '돌아오/V... |
| 202049 | 2020-12-31 00:00:00 | UP | <딜링룸 백브리핑> NFL 선수 오름, 비트코인으로 연봉 수령 | (서울=연합인포맥스) 국제경제부 = 미국 프로축구(NFL) 캐롤라이나 팬서스팀의 공... | ['국제/NNG', '경제부/NNG', '프로/NNG', '축구/NNG', '스팀/... |
| 202050 | 2020-12-31 00:00:00 | UP | <뉴욕은 지금> 브렉시트와 디스플레이션의 수출 | (뉴욕=연합인포맥스) 2020.12.24 vs 2020.12.30 의미 없는 숫자의... | ['뉴욕/NNG', '의미/NNG', '없/VV', '숫자/NNG', '나열/NNG... |

전처리 과정

For Call Rate

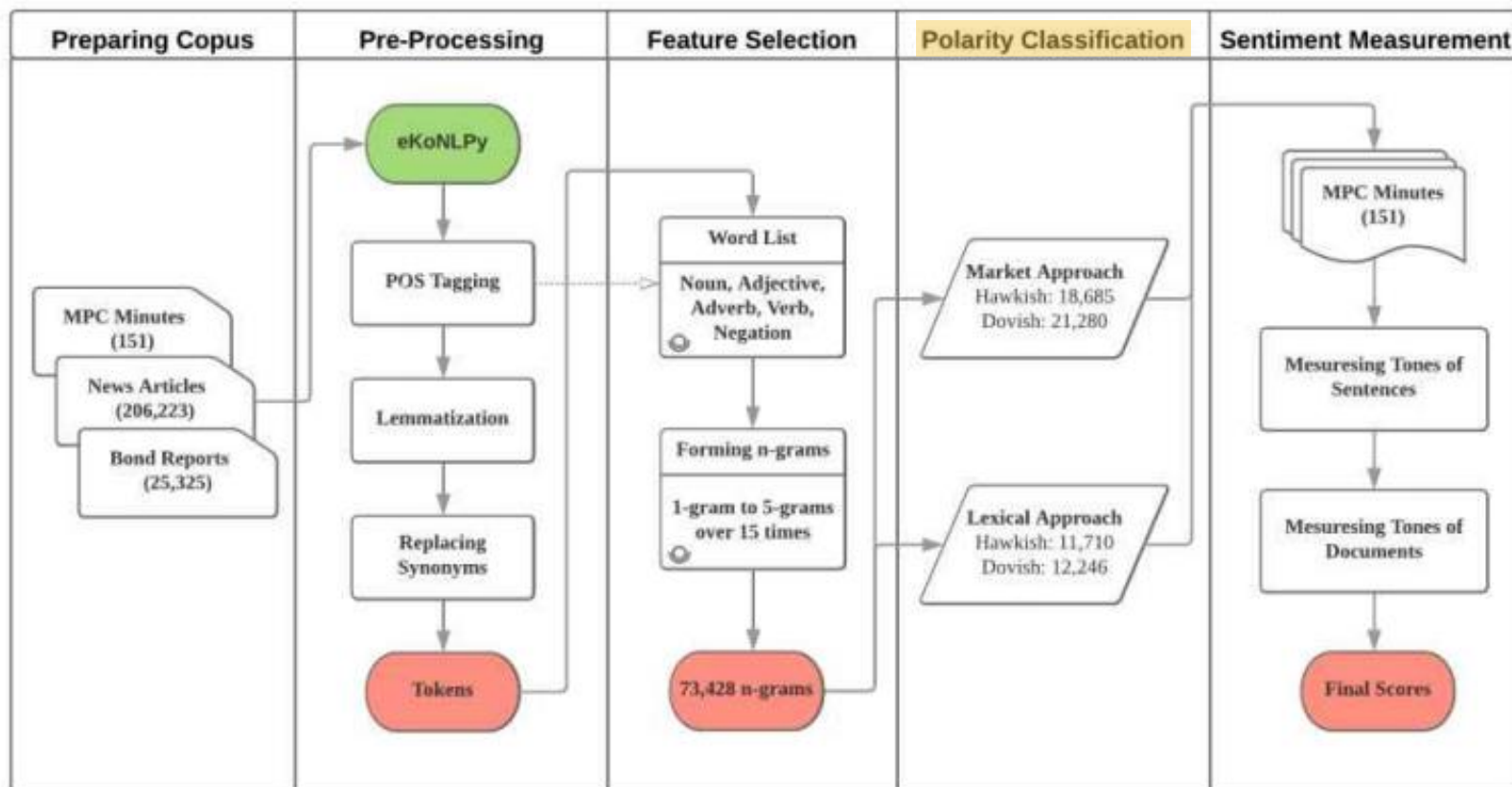
```
1 import requests
2 from bs4 import BeautifulSoup
3 import re
4 f = open("call.csv", "w")
5 f.write("날짜,콜금리,₩n")
6 for i in range(1,400):
7     url = 'https://finance.naver.com/marketindex/interestDailyQuote.nhn?marketindexCd=IRR_CALL&page='+str(i)
8     response = requests.get(url)
9     soup = BeautifulSoup(response.text, 'html.parser')
10    list_call = soup.select('div[class="section_exchange2"] tbody > tr')
11    for call in list_call:
12        date=str(call.select('td')[0].text.strip())
13        call_index=str(call.select('td')[1].text.strip())
14        f.write(str(date)+","+str(call_index)+'₩n')
15 f.close()
```

| | date | callrate | before_date | before_callrate | updown |
|------|------------|----------|-------------|-----------------|--------|
| 0 | 2011.01.03 | 2.49 | 2010.12.01 | 2.50 | DOWN |
| 1 | 2011.01.04 | 2.50 | 2010.12.02 | 2.50 | MIDDLE |
| 2 | 2011.01.05 | 2.49 | 2010.12.03 | 2.50 | DOWN |
| 3 | 2011.01.06 | 2.49 | 2010.12.06 | 2.50 | DOWN |
| 4 | 2011.01.07 | 2.49 | 2010.12.07 | 2.52 | DOWN |
| ... | ... | ... | ... | ... | ... |
| 2467 | 2020.12.29 | 0.54 | 2020.11.25 | 0.51 | UP |
| 2468 | 2020.12.30 | 0.57 | 2020.11.26 | 0.53 | UP |
| 2469 | 2020.12.31 | 0.61 | 2020.11.27 | 0.52 | UP |
| 2470 | 2021.01.04 | 0.53 | 2020.11.30 | 0.60 | DOWN |
| 2471 | 2021.01.05 | 0.48 | 2020.12.01 | 0.51 | DOWN |

전처리 과정

| | date | updown | title | content | ngrams |
|--------|---------------------|--------|--|--|---|
| 0 | 2011-01-03 00:00:00 | DOWN | [마켓in]새해 외화채 발행 위축될듯..“만기”-조달금리↑ | // flash 오류를 우회하기 위한 함수 추가\nfunction _flash_re... | ['외표채/NNG', '만기도래액/NNG', '추정/NNG', '사기업/NNG', ... |
| 1 | 2011-01-03 00:00:00 | DOWN | [마켓in][채권물]"금리 되돌림..캐리수요 유입 관심" | // flash 오류를 우회하기 위한 함수 추가\nfunction _flash_re... | ['국고채/NNG', '발행/NNG', '불확실성/NNG', '해소/NNG', '되... |
| 2 | 2011-01-03 00:00:00 | DOWN | [마켓in]동부證 "3년물리 금리상승 견인" | // flash 오류를 우회하기 위한 함수 추가\nfunction _flash_re... | ['시장/NNG', '기사/NNG', '프리미엄/NNG', '시장/NNG', '출고... |
| 3 | 2011-01-03 00:00:00 | DOWN | "5% 고정장" 방점..한은 기준금리 정상화 '힘'로 | // flash 오류를 우회하기 위한 함수 추가\nfunction _flash_re... | ['성장/NNG', '없이/MAG', '일자리/NNG', '복지/NNG', '재정/... |
| 4 | 2011-01-03 00:00:00 | DOWN | [마켓in]하나대투證 "올해 채권금리 상승..국고3년 4%대" | // flash 오류를 우회하기 위한 함수 추가\nfunction _flash_re... | ['시장/NNG', '기사/NNG', '프리미엄/NNG', '시장/NNG', '출고... |
| ... | ... | ... | ... | ... | ... |
| 202046 | 2020-12-31 00:00:00 | UP | 연말 원화 자금시장 안정의 '숨은 공'...韓銀 5조원 풀었다 | (서울=연합인포맥스) 노현우 기자 = 단기 자금시장이 연말에도 안정세를 보이는 가운데... | ['기자/NNG', '단기자금시장/NNG', '안정/NNG', '보/VV', '가운... |
| 202047 | 2020-12-31 00:00:00 | UP | "올해 세계 상장기업 시가총액 1경6천300조원 증가...중앙은행 완화 덕" | (서울=연합인포맥스) 문정현 기자 = 올해 세계 증시가 상승하면서 상장 기업의 시가... | ['기자/NNG', '세계/NNG', '증시/NNG', '상승/NNG', '하/XS... |
| 202048 | 2020-12-31 00:00:00 | UP | 비앙코 "2021년, 30년만에 물가 돌아오는 해 될 것" | (뉴욕=연합인포맥스) 우성문 통신원 = 비앙코 리서치의 짐 비앙코 대표가 2021년... | ['뉴욕/NNG', '짐/NNG', '물가/NNG', '다시/MAG', '돌아오/V... |
| 202049 | 2020-12-31 00:00:00 | UP | <딜링룸 백브리핑> NFL 선수 오쿤, 비트코인으로 연봉 수령 | (서울=연합인포맥스) 국제경제부 = 미국 프로축구(NFL) 캐롤라이나 팬서스팀의 공... | ['국제/NNG', '경제부/NNG', '프로/NNG', '축구/NNG', '스팀/... |
| 202050 | 2020-12-31 00:00:00 | UP | <뉴욕은 지금> 브렉시트와 디플레이션의 수출 | (뉴욕=연합인포맥스) 2020.12.24 vs 2020.12.30 의미 없는 숫자의... | ['뉴욕/NNG', '의미/NNG', '업/VA', '숫자/NNG', '나열/NNG... |

Figure 1. Procedure of Sentiment Analysis



Note: This figure, along with table 1, summarizes our discussion in section 3.

사전구축

202051 rows × 5 columns

```
1 total_ngrams = []
2 specific_ngrams = []
3 for i in range(100001, 150001):
4     print(i)
5     if merge_file['ngrams'][i] == 'nan' or merge_file['ngrams'][i] == 'None' or type(merge_file['ngrams'][i]) == float:
6         pass
7     else:
8         for j in range(len(merge_file['ngrams'][i][1:-1].split(','))):
9             try:
10                 total_ngrams.append(re.findall(r'(?<=\\#').+(?=\\#)', merge_file['ngrams'][i][1:-1].split(',')[j])[0])
11             except:
12                 pass
13 for v in total_ngrams:
14     if v not in specific_ngrams:
15         specific_ngrams.append(v)
```

```
1 len(specific_ngrams)
```

82485

```
1 len(total_ngrams)
```

9887377

사전구축

```
1  #업 다운 카운트
2  list_count = []
3  order = 0
4  for ngram in specific_ngrams:
5      print(order)
6      up_count = 0
7      down_count = 0
8      for i in range(len(list_ngrams)):
9          if ngram not in list_ngrams[i]:
10             pass
11          else:
12             if merge_file['updown'][i+100001] == "UP":
13                 up_count += list_ngrams[i].count(ngram)
14             elif merge_file['updown'][i+100001] == "DOWN":
15                 down_count += list_ngrams[i].count(ngram)
16      count = [ngram, up_count, down_count]
17      list_count.append(count)
18      order += 1
```

사전구축

$$\text{polarity score} = \frac{p(\text{feature}|\text{hawkish})}{p(\text{feature}|\text{dovish})} = \frac{p(\text{feature}\&\text{hawkish})/p(\text{hawkish})}{p(\text{feature}\&\text{dovish})/p(\text{dovish})} \quad (1)$$

```
#polarity score 부여
up_sum = 0
for i in range(len(ngram_dictionary)):
    up_sum += ngram_dictionary['up_count'][i]

down_sum = 0
for j in range(len(ngram_dictionary)):
    down_sum += ngram_dictionary['down_count'][i]

for i in range(len(ngram_dictionary)):
    ngram_dictionary['polarity score'][i] = (ngram_dictionary['up_count'][i]/up_sum)/(ngram_dic
```

```
#polarity 분류
for i in range(len(ngram_dictionary)):
    if ngram_dictionary['polarity score'][i] > 1.3:
        ngram_dictionary['polarity'][i] = 'Hawkish'
    elif ngram_dictionary['polarity score'][i] < 0.76:
        ngram_dictionary['polarity'][i] = 'dovish'
    else:
        ngram_dictionary['polarity'][i] = 'nothing'
```

사전구축

```
1 pd.read_csv('Hawkish_dic.csv' , encoding = 'utf-8')
```

| | index | ngram | total_count | up_count | down_count | polarity score | polarity |
|------|-------|------------------------|-------------|----------|------------|----------------|----------|
| 0 | 424 | 긴축/NNG | 12684 | 6681 | 6003 | 1.394243 | Hawkish |
| 1 | 544 | 저축은행/NNG | 9412 | 4937 | 4475 | 1.382088 | Hawkish |
| 2 | 679 | 절하/NNG | 7362 | 4073 | 3289 | 1.551372 | Hawkish |
| 3 | 690 | 프리미엄/NNG | 7225 | 4117 | 3108 | 1.659454 | Hawkish |
| 4 | 810 | 중립/NNG | 5855 | 2991 | 2864 | 1.308305 | Hawkish |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 8958 | 36990 | 우려/NNG;시장/NNG;불확실성/NNG | 16 | 13 | 3 | 5.428597 | Hawkish |
| 8959 | 36997 | 내림/NNG;전환/NNG | 16 | 14 | 2 | 8.769272 | Hawkish |
| 8960 | 37008 | 신건/NNG | 16 | 14 | 2 | 8.769272 | Hawkish |
| 8961 | 37010 | 활태/NNG | 16 | 16 | 0 | inf | Hawkish |
| 8962 | 37013 | 건전성/NNG;부담/NNG | 16 | 9 | 7 | 1.610683 | Hawkish |

8963 rows × 7 columns

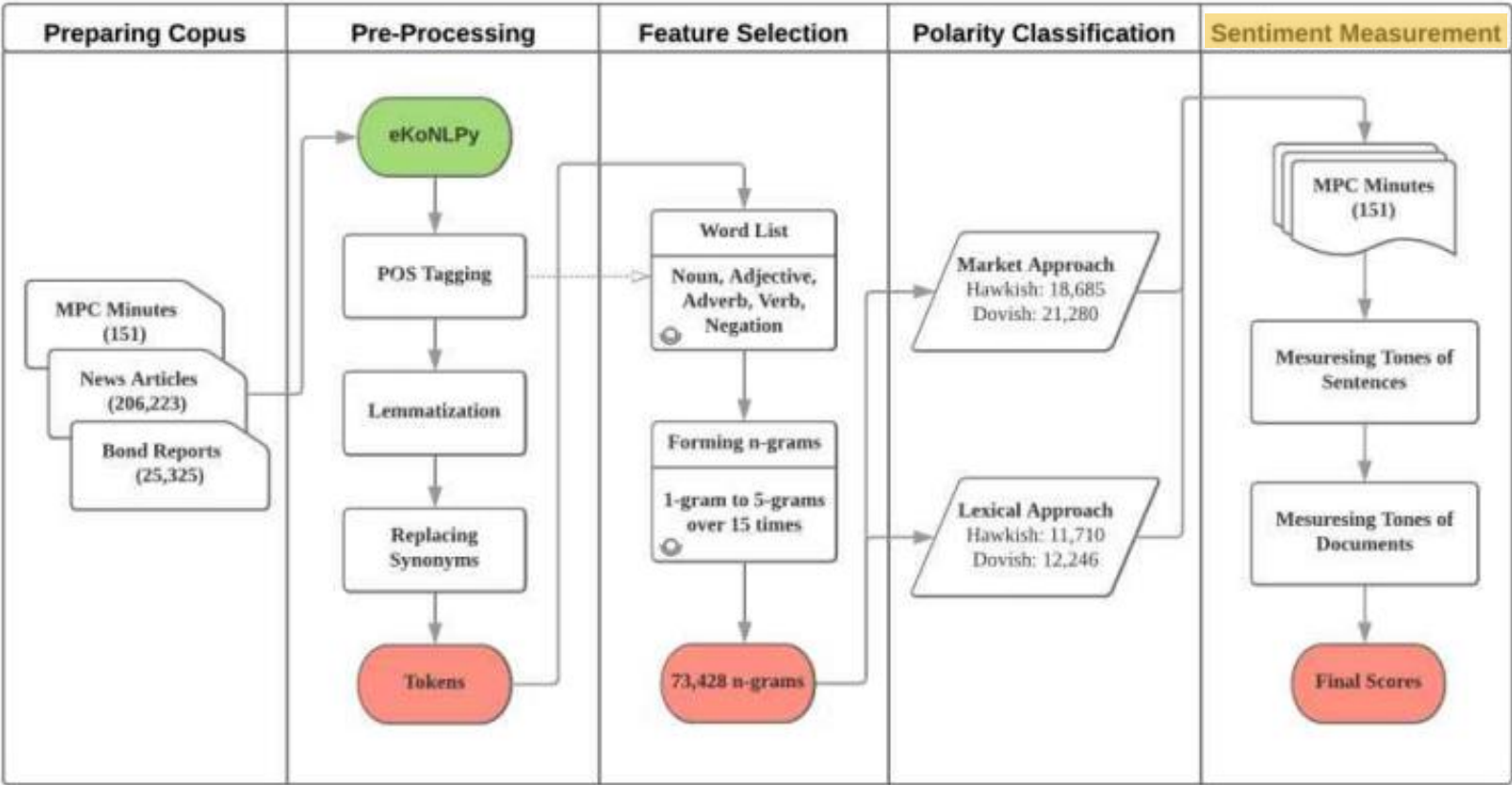
```
1 pd.read_csv('dovish_dic.csv' , encoding = 'utf-8')
```

| | index | ngram | total_count | up_count | down_count | polarity score | polarity |
|------|-------|-----------------------|-------------|----------|------------|----------------|----------|
| 0 | 25 | 인하/NNG | 102562 | 36957 | 65605 | 0.705708 | Dovish |
| 1 | 115 | 코로나/NNG | 43259 | 14963 | 28296 | 0.662459 | Dovish |
| 2 | 136 | 금리/NNG;인하/NNG | 37580 | 12930 | 24650 | 0.657124 | Dovish |
| 3 | 378 | 브렉시트/NNG | 14669 | 3835 | 10834 | 0.443447 | Dovish |
| 4 | 486 | 바이러스/NNG | 10713 | 3421 | 7292 | 0.587722 | Dovish |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 8995 | 37003 | 권선/NNG | 16 | 4 | 12 | 0.417584 | Dovish |
| 8996 | 37004 | 확충안/NNG | 16 | 6 | 10 | 0.751652 | Dovish |
| 8997 | 37005 | 개선/NNG;속도/NNG;둔화/NNG | 16 | 3 | 13 | 0.289097 | Dovish |
| 8998 | 37007 | 합/MAG | 16 | 2 | 14 | 0.178965 | Dovish |
| 8999 | 37012 | 노동/NNG;생산성/NNG;개선/NNG | 16 | 5 | 11 | 0.569433 | Dovish |

9000 rows × 7 columns

Total_count가 15개 이하인 값들은 삭제.

Figure 1. Procedure of Sentiment Analysis



Note: This figure, along with table 1, summarizes our discussion in section 3.

의사록 논조파악

사전이용 금통위의사록의 결과값을 분석

$$tone_s = \frac{\text{No. of hawkish features} - \text{No. of dovish features}}{\text{No. of hawkish features} + \text{No. of dovish features}}$$

$$tone_i = \frac{\text{No. of hawkish } tone_{s,i} - \text{No. of dovish } tone_{s,i}}{\text{No. of hawkish } tone_{s,i} + \text{No. of dovish } tone_{s,i}}$$

의사록 논조파악

```
#어조 분석
tone_list = [] #각 의사록별 톤 리스트 도출
for i in range(len(mp_total)): #각 의사록
    negative_sentence = 0
    positive_sentence = 0

    for j in range(len(mp_total[i])): #각 문장 -> 문장의 sentence_tone을 도출
        negative_word = 0
        positive_word = 0
        for k in range(len(mp_total[i][j])): #각 문장의 단어들
            if mp_total[i][j][k] in negative_list:
                negative_word += 1
            elif mp_total[i][j][k] in positive_list:
                positive_word += 1
            else:
                pass
        try:
            sentence_tone = (positive_word - negative_word) / (positive_word + negative_word)
        except:
            sentence_tone = 0

        if sentence_tone > 0:
            positive_sentence += 1
        elif sentence_tone < 0:
            negative_sentence += 1
        else:
            pass

    try:
        document_tone = (positive_sentence - negative_sentence) / (positive_sentence + negative_sentence)
        tone_list.append(document_tone)
    except:
        document_tone = 0 #각 의사록별 document_tone을 도출
        tone_list.append(document_tone)
```

결과값 비교

| | date | TONE-Doc |
|-----|------------|-----------|
| 0 | 2011.01.13 | 0.417219 |
| 1 | 2011.02.11 | 0.777778 |
| 2 | 2011.03.10 | 0.619048 |
| 3 | 2011.04.12 | 0.626866 |
| 4 | 2011.05.13 | 0.250000 |
| ... | ... | ... |
| 96 | 2020.05.28 | -0.818182 |
| 97 | 2020.07.16 | -0.142857 |
| 98 | 2020.08.27 | -0.371429 |
| 99 | 2020.10.14 | -0.187500 |
| 100 | 2020.11.26 | 0.000000 |

| | date | TONE-Doc | baserate |
|-----|------------|-----------|----------|
| 0 | 2011.01.13 | 0.417219 | 2.75 |
| 1 | 2011.02.11 | 0.777778 | 2.75 |
| 2 | 2011.03.10 | 0.619048 | 3.00 |
| 3 | 2011.04.12 | 0.626866 | 3.00 |
| 4 | 2011.05.13 | 0.250000 | 3.00 |
| ... | ... | ... | ... |
| 96 | 2020.05.28 | -0.818182 | 0.50 |
| 97 | 2020.07.16 | -0.142857 | 0.50 |
| 98 | 2020.08.27 | -0.371429 | 0.50 |
| 99 | 2020.10.14 | -0.187500 | 0.50 |
| 100 | 2020.11.26 | 0.000000 | 0.50 |

101 rows x 3 columns

```
corr = mp_tone.corr()
print(corr)
```

| | TONE-Doc | baserate |
|----------|----------|----------|
| TONE-Doc | 1.000000 | 0.173368 |
| baserate | 0.173368 | 1.000000 |



| | date | TONE-Doc | baserate |
|-----|------------|-----------|----------|
| 0 | 2011.01.13 | 0.366197 | 2.75 |
| 1 | 2011.02.11 | 0.000000 | 2.75 |
| 2 | 2011.03.10 | 0.460993 | 3.00 |
| 3 | 2011.04.12 | 0.255814 | 3.00 |
| 4 | 2011.05.13 | 0.200000 | 3.00 |
| ... | ... | ... | ... |
| 96 | 2020.05.28 | -0.729730 | 0.50 |
| 97 | 2020.07.16 | -0.571429 | 0.50 |
| 98 | 2020.08.27 | -0.636364 | 0.50 |
| 99 | 2020.10.14 | -0.466667 | 0.50 |
| 100 | 2020.11.26 | -0.272727 | 0.50 |

101 rows x 3 columns

```
corr = mp_tone2.corr()
print(corr)
```

| | TONE-Doc | baserate |
|----------|----------|----------|
| TONE-Doc | 1.0000 | 0.1202 |
| baserate | 0.1202 | 1.0000 |

결과값 비교



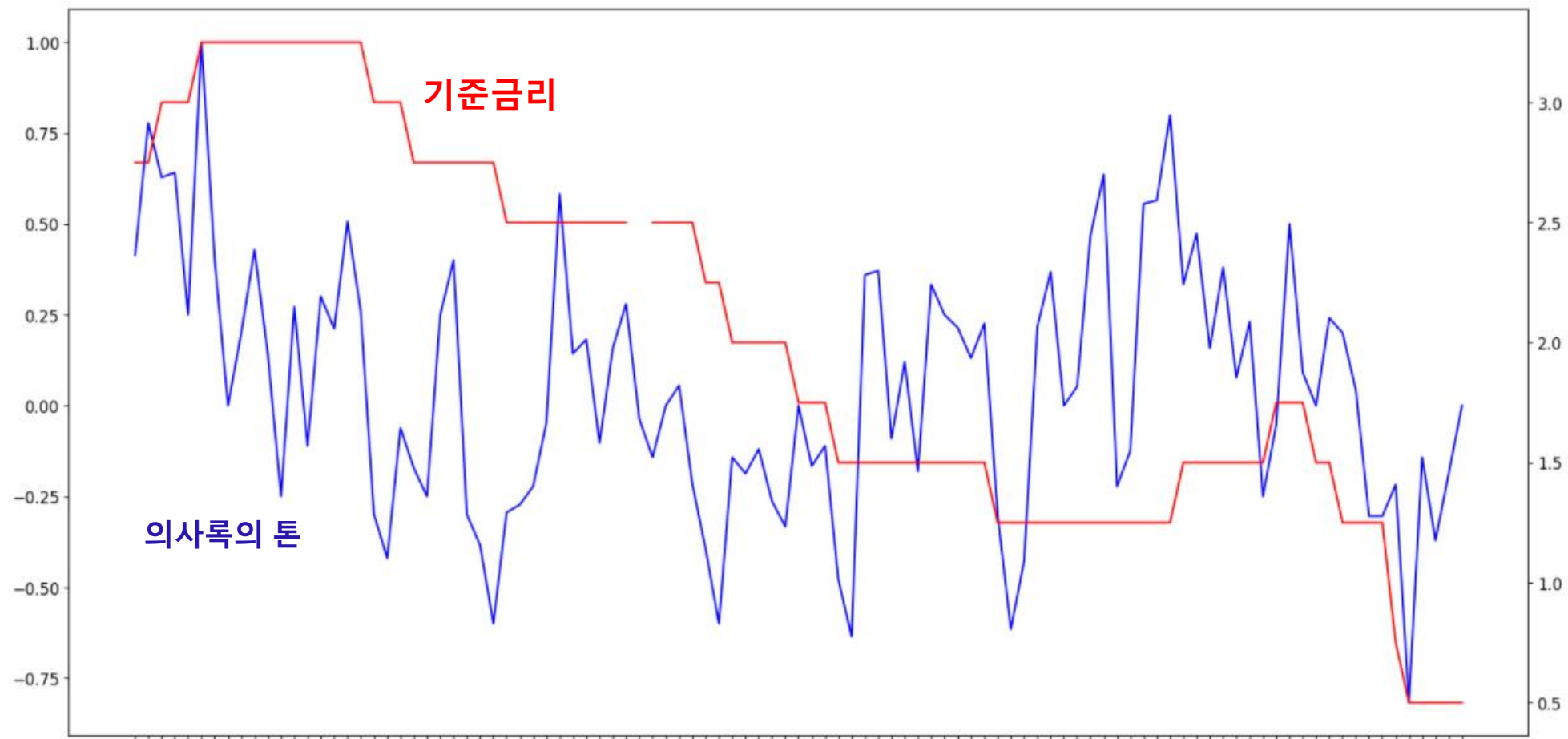
<부정사전>

결과값 비교

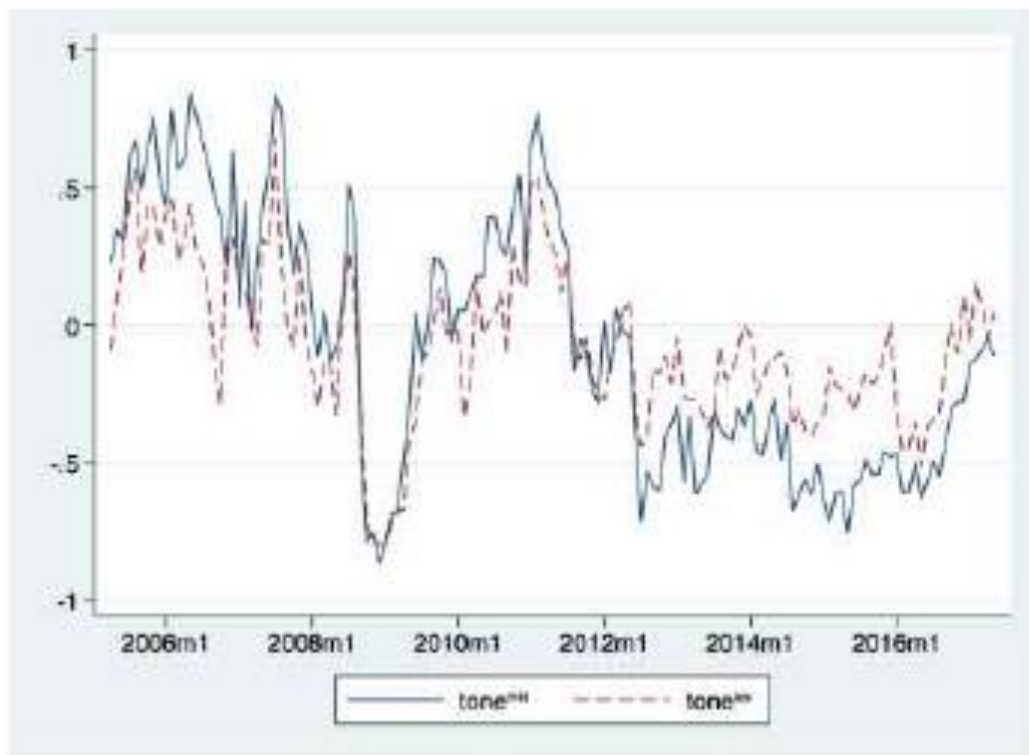


<긍정사전>

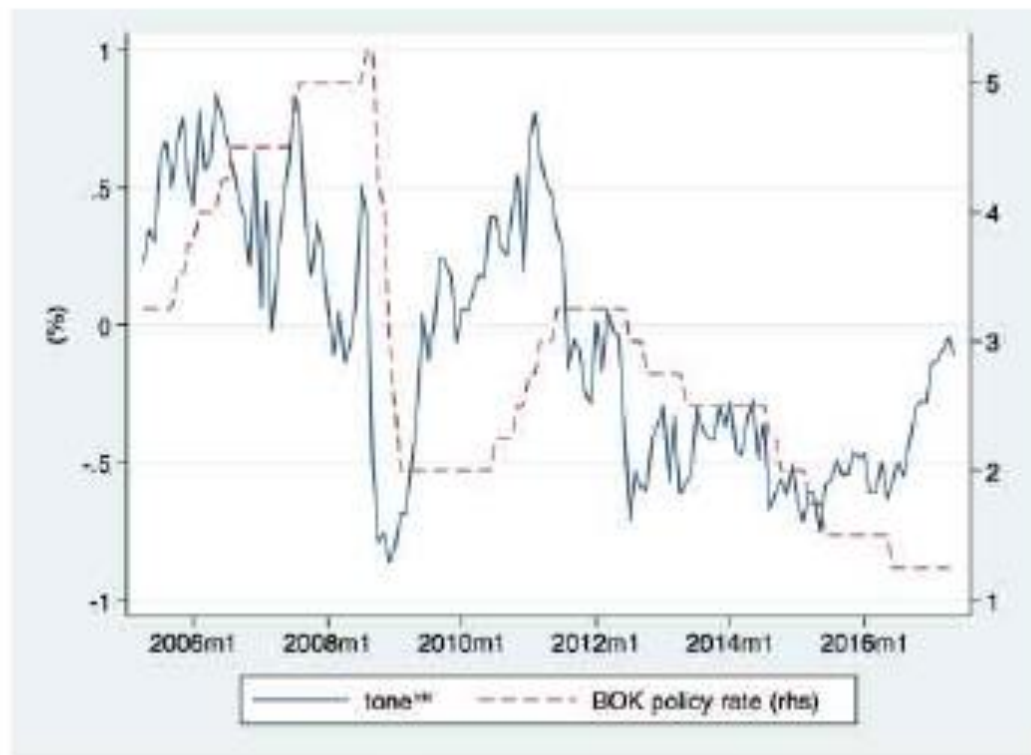
결과값 비교



결과값 비교

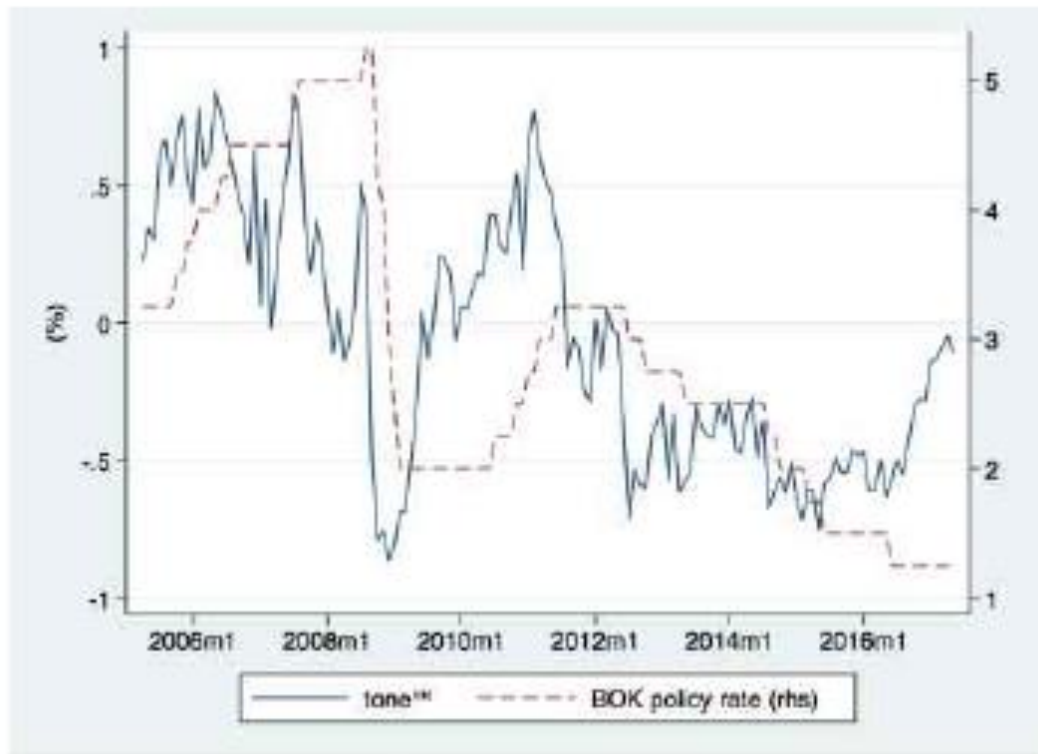


(a) tone^{mkt} and tone^{leiz}

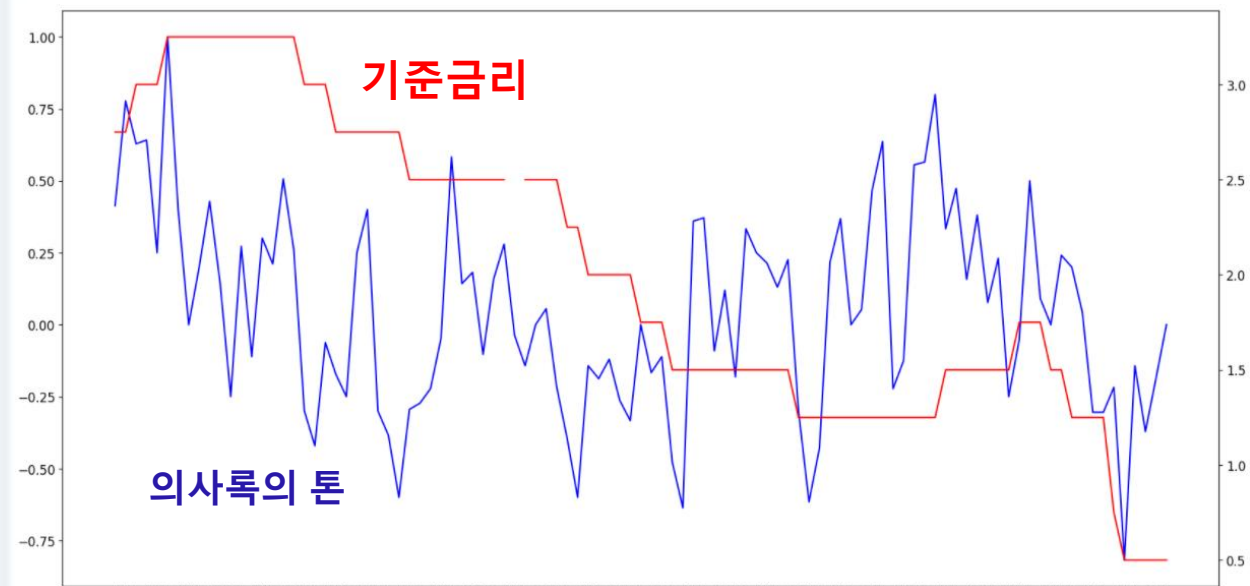


(b) tone^{mkt} and BOK policy rate

결과값 비교



(b) $tone^{mkt}$ and BOK policy rate



결과값 비교

1. 단기 변동에 대한 설명력이 장기간 변동보다 높다.

장기 : 2011~2020 상관계수 0.17



단기 : 2011~2015 상관계수 0.44

단기 : 2016~2020 상관계수 0.312.

2. 어조는 변동폭이 크고, 기준금리 및 콜금리는 변동폭이 작다.

어조의 경우 경기 및 이벤트에 따라 변동폭 크게 변화.
기준금리는 글로벌금융위기 이후 완만한 하락세.

3. 어조가 기준금리를 약하게 선행한다.

금통위 어조와 '1년 후 기준금리'의 상관계수 '0.17' -> '0.22' 소폭 상승.
2011~2015 : '0.44' -> '0.6'

금통위 어조가 기준금리에 어느정도 영향. 하지만 시장상황 변화로
추가적인 연구가 필요.

감사합니다