텍스트 마이닝을 활용한 통화정책 분석

Introduction to eKoNLPy
(Korean NLP python package for Economic Analysis)

Youngjoon Lee (yj.lee@yonsei.ac.kr)

July 16, 2018

Yonsei School of Business
(https://github.com/entelecheia/eKoNLPy)

Table of contents

1. Introduction

2. Building the Monetary Policy Sentiment Lexicon

3. Monetary Sentiment Analysis

4. Comparison with English Sentiment Aanalysis

5. Topic Sentiment Analysis

6. Appendix: eKoNLPy Guide

Introduction

Introduction

연구목적

- 문서에서 통화정책의 비정량적 정보를 측정하고 그 효과를 검증하고자 함
- 통화정책 센티멘트(톤) 측정을 위한 한국어 사전 구축
- 경제/경영 분야에서 한국어 텍스트 마이닝 활용 방안 제시
- 다양한 후속 연구를 지원하기 위한 분석 툴 제공

연구현황 및 문제점

- 해외의 경우 경제/경영 분야에서 텍스트 마이닝을 활용한 다양한 연구가 이루어 지고 있음
- 한국의 경우도 다른 분야에서는 활발한 연구가 이루어 지고 있으나 경제/경영 분야는 초기 단계임
- 경제/경영 분야의 특수성으로 인해 분석을 지원하는 툴이 제한적이기 때문
- 현존하는 많이 형태소 분석기는 경제/경영 분야 전문용어, 외래어, 외국어 처리에 미흡
- 경제 분야 감성분석을 위한 사전 부재
- 경제/경영 분야 연구자들이 손쉽게 사용할 수 있는 툴 부재

Introduction

공헌점

- 통화정책 센티멘트 측정을 위한 최초의 한국어 사전 구축
- 경제학 분석에서 텍스트 마이닝 기법의 다양한 활용 방안 제시
- 경제/경영 분야에서 텍스트 마이닝 분석을 할 경우, 전문용어, 외래어, 외국어 등으로 인해 형태소 분석이 어려운 문제점 해결
- 후속 연구를 위한 분석 툴 제공

향후 연구주제

- 금통위 전후 시장의 기대 변화를 텍스트에서 직접적으로 측정하여 통화정책 효과 분석
- 통화정책 기대변화가 자산 가격에 미치는 영향
- 기업실적 센티멘트와 통화정책 센티멘트를 측정하고 이들이 자산가격과 자본구조 등에 미치는 영향

Building the Monetary Policy Sentiment Lexicon

통화정책 사전구축 절차

1. 통화정책 관련 텍스트 수집
   - 뉴스, 애널리스트 분석리포트, 금통위 의사록 등 통화정책 관련 텍스트 자료 수집
2. 전처리 및 형태소 분석
   - 문서에서 텍스트 부분만을 추출하여 개별 문장으로 구분한 후, 형태소 분석 실시
3. 동의어 처리 및 Lemmatisation
   - 효율적인 학습을 위해 같은 의미를 가지고 있으나 다양한 변형이 발생하는 경우에 이를 통일하는 과정
   - 동의어, 약어, 외래어 등을 대표 단어로 통일
   - 형용사, 동사, 부사 등 변형이 일어나는 형태소의 경우에 표제어를 통일하는 작업
4. n-gram 단어 사전 구축
   - 개별 단어로는 통화정책 방향에 대한 판단이 어렵기 때문에 연속된 n개의 단어의 조합을 사용
   - 명사, 동사, 부사, 부정어 등을 포함하는 5-gram 단어 사전 구축
5. Polarity lexicon 생성
   - Market approach: 문서가 공개된 시점의 시장 변수의 움직임을 통해 단어의 극성 구분
   - Lexical approach: 단어간 관계에 기초하여 단어의 극성을 구분
   - Supervised learning approach: 전문가에 의해 극성이 분류된 문서에 대한 학습을 통해 극성 분류기 (classifier) 구축

절차 1: 텍스트 수집

유효한 사전구축을 위해서 통화정책과 관련된 다량의 문서를 수집

1. 금통위 의사록
   - 한국은행의 공식적인 금통위 회의 결과 자료
2. 금리 관련 뉴스
   - 시장이 받아들이는 통화정책 관련 정보를 확인할 수 있는 문서
3. 채권 애널리스트 분석보고서
   - 통화정책 전망과 해석을 살펴볼 수 있는 증권사 채권전문가의 분석보고서

문서 구분	문서수	평균 문장수	최대 문장수
금통위 의사록	151	165	326
뉴스	206,223	15	340
분석보고서	26,284	49	2,515
전체	232,658	19	2,515

절차 1: 텍스트 수집 - 금통위 의사록

- 공개 시점
  - 2005년 4월 이후: 6주후 공개
  - 2012년 9월 이후: 회의일로부터 2주가 경과한 이후 최초로 도래하는 화요일에 공개
  - 발표시각: 오후 4시 발표, 장중 금융시장 반응 확인 불가
- 의사록의 구성
  - 행정일반정보
  - 토의내용: 국내외 경제동향, 외환·금융시장 동향, 통화정책 방향
  - (정부측 열석자 발언)
  - 한국은행 기준금리 결정에 관한 위원별 의견개진
  - 토의결론
  - 심의결과

절차 1: 텍스트 수집 - 금통위 의사록

- 의사록 수집대상 기간: 2005.5 - 2017.12

Figure 1: 의사록 문장수 추이

Figure 2: 의사록 영역별 문장수 추이

절차 1: 텍스트 수집 - 뉴스

- 분석대상 신문사: 연합인포맥스, 이데일리, 연합뉴스 (보도비중 상위 3개사)
- 분석대상 신문사의 뉴스 중에서 본문에 '금리'를 포함하는 경제분야 뉴스 전부를 수집
- 수집대상 기간: 2005.1 - 2017.12

Figure 3: 월별 뉴스 발생 추이

Figure 4: 년도별 뉴스 발생 추이

절차 1: 텍스트 수집 - 채권 분석보고서

- 증권사 채권 애널리스트가 발행하는 분석보고서
- 채권투자전략, 금리전망, 채권시황 등의 내용을 포함
- 수집대상 기간: 2004.1 ~ 2017.12

Figure 5: 월별 채권 분석리포트 발간 추이

절차 2: 전처리 및 형태소 분석

- 텍스트 마이닝을 위해 문서를 문장으로 나누고, 문장을 단어로 나누는 작업이 필요 (Tokenizing)
- 한글의 경우 단어가 띄어쓰기로 구분되지 않는 언어의 특성상 형태소 분석은 필수
- 많이 사용하는 다수의 한글 형태소 분석기가 존재하나 경제/경영 분야 분석에는 적합하지 않음
- 일반적으로 사용하지 않는 전문용어, 외래어, 외국어 등에 대한 형태소 분석이 제대로 이루어지지 않음
- 경제/경영 분야 용어사전을 사용하고 형태소 분해 기준 등을 수정한 형태소 분석기 직접 제작함

절차 2: 전처리 및 형태소 분석

"한국은행이 12일 금융통화위원회(금통위) 회의를 열고 기준금리를 현행 연 1.50%로 동결했다."

- 형태소 분석결과 비교
  - 기존 형태소 분석기 (KoNLPy) 사용 결과:
    ['한국은행/NNP', '이/JKS', '12/SN', '일/NNBC', '금융/NNG', '통화/NNG', '위원회/NNG', '(/SSO', '금/NNG', '통/NNG', '위/NNG', ')/SSC', '회의/NNG', '를/JKO', '열/VV', '고/EC', '기준/NNG', '금리/NNG', '를/JKO', '현행/NNG', '연/NNG', '1/SN', './SY', '50/SN', '%/SY', '로/JKB', '동결/NNG', '했/XSV', '다/EF', './SF']
  - 자체 개발 형태소 분석기 (eKoNLPy) 사용 결과:
    ['한국은행/NNG', '이/JKS', '공공요금/NNG', '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', './SY', '50/SN', '%/SY', '로/JKB', '동결/NNG', '했/XSV', '다/EC']
  - 기존 형태소 분석기는 금융통화위원회/금통위 단어를 인식하지 못함

절차 2: 전처리 및 형태소 분석

"금리 박스권 상단 상향과 일드 커브 완만한 스티프닝 전망"

- 형태소 분석결과 비교
  - 기존 형태소 분석기 (KoNLPy) 사용 결과:
    ['금리/NNG', '박스/NNG', '권/XSN', '상단/NNG', '상향/NNG', '과/JC', '일/NNG', '드/NNG', '커브/NNG', '완만/XR', '한/XSA', '스티/NNP', '프/VX', '닝/NNP', '전망/NNG']
  - 자체 개발 형태소 분석기 (eKoNLPy) 사용 결과:
    ['금리/NNG', '박스권/NNG', '상단/NNG', '상향/NNG', '과/JC', '일드커브/NNG', '완만/NNG', '한/XSA', '스티프닝/NNG', '전망/NNG']
  - 기존 형태소 분석기는 '일드커브', '대손충당금비율', '스티프닝' 이란 단어를 인식하지 못함

절차 3: 동의어 처리 및 Lemmatisation

- 효율적인 분석과 학습을 위해 같은 의미를 가지고 있으나 다양한 변형이 발생하는 경우에 이를 통일하는 과정이 필요
- 동의어 처리
  - 유사어, 약어, 외래어 등을 대표 단어로 통일
  - eKoNLPy에서 1,326개 경제분야 단어의 동의어 처리
  - ex) ['리스크': '위험'], ['스와프': '스왑'], ['스탠스': '기조'], ['인플레': '인플레이션'], ['外人': '외국인'] 등
- Lemmatisation
  - 형용사, 동사 등 변형이 일어나는 형태소의 경우에 표제어를 통일하는 작업
  - eKoNLPy에서 1,287개 단어의 lemma 처리
  - ex) ['오른', '오른다고', '오른다는', '올라', '올라서'] 등의 경우 표제어 ['오르']로 처리

절차 4: n-gram 단어 사전 구축

- 개별 단어로는 통화정책 방향에 대한 판단이 어렵기 때문에 연속된 n개의 단어의 조합을 사용
- 명사, 동사, 부사 및 부정어 등을 포함하는 5-gram을 사용
- 발생빈도수 15개 이상인 73,428개의 5-gram을 분석대상으로 사용

Table 1: n-gram 예시

금리/NNG;지준율/NNG;인하/NNG	총액/NNG;대출/NNG;한도/NNG;축소/NNG
하락/NNG;거래/NNG;마치/VV;내리/VV	재할인율/NNG;인상/NNG
금리/NNG;인하/NNG;가계/NNG;부채/NNG;증가/NNG	인플레이션/NNG;심리/NNG;확산/NNG
금리/NNG;인하/NNG;소식/NNG;상승/NNG	자본/NNG;유출입/NNG;변동/NNG;완화/NNG
정책/NNG;공조/NNG;금리/NNG;인하/NNG	자본/NNG;유출입/NNG;규제/NNG;우려/NNG
실물/NNG;경기/NNG;침체/NNG;우려/NNG	물가/NNG;안정/NNG;견조/NNG;성장/NNG
유동성/NNG;경색/NNG;해소/NNG	금리/NNG;인상/NNG;인플레이션/NNG;우려/NNG
경제주체/NNG;심리/NNG;부진/NNG	소비자/NNG;물가/NNG;상승률/NNG;금리/NNG;인상/NNG
금융위기/NNG;세계/NNG;확산/NNG	물가/NNG;불안/NNG;확산/NNG
구조조정/NNG;자본/NNG;확충/NNG	잠재/NNG;성장률/NNG;경제/NNG;성장/NNG

절차 5: Polarity Lexicon 생성

5-gram 단어의 극성(polarity)을 구분하여 Positive/Negative를 나타내는 lexicon을 생성하는 절차

1. Market approach
   - 문서가 공개된 시점의 시장 변수의 움직임을 통해 단어의 극성 구분
   - 전문가의 판단, 대중의 평가 등의 주관적 판단을 배제하고 시장에 내재되어 있는 정보를 바탕으로 단어의 극성을 판단할 수 있는 방법
   - 넓은 범위의 단어와 통계적 유의성 확보를 위해 대량의 문서 데이터에 적합
2. Lexical approach
   - 단어간 관계 네트워크 그래프에서 Seed word를 사용해 label propagation을 적용하는 방식
   - 단어 네트워크 구축 방식에 따라 크게 코퍼스기반(Corpus-based) 접근법과 사전기반(Dictionary-based) 접근법이 존재
3. Supervised learning approach
   - 전문가에 의해 극성이 분류된 문서에 대한 학습을 통해 극성 분류기(classifier) 구축
   - 감독기반 학습법으로 비교적 적은 문서를 가지고 우수한 분류 능력을 확보할 수 있음
   - 소량의 대표 샘플에 기초하기 때문에 넓은 범위의 문서에 적용에는 한계 존재

절차 5: Polarity Lexicon 생성 - Market approach

1. 회귀분석
   - Jegadeesh and Wu (2013)[6] 는 Harvard IV-4와 Loughran and McDonald (2011)[8] 사전의 단어를 설명변수로 하고 동기간 주가수익률을 종속변수로 하는 회귀분석을 통해 얻어진 계수를 각 단어의 극성을 나타내는 점수로 사용
   - Logistic regression 등도 사용 가능
2. SO-PMI (Semantic Orientation from Point-wise Mutual Information)
   - PMI는 원래 단어 사이의 연관성을 측정하는 지표로 사용 (Church and Hanks 1989)[2]
   - Positive/Negative 시장반응과 단어간의 연관성을 측정하면 시장반응으로 단어의 극성 분류 가능

$$SO\text{-}PMI(word) = PMI(word, positive\ label) - PMI(word, negative\ label)$$

$$PMI(word, label) = \log \frac{p(word, label)}{p(word)p(label)}$$

3. Naive-bayes classifier
   - Bayes의 조건부 확률을 사용하여 분류(classification) 오류를 최소화하는 기계학습 (Machine-learning) 알고리즘으로 스팸메일이나 문서 종류의 구분과 같이 다양한 곳에 사용됨
   - 주어진 클래스의 한 속성(feature) 값이 다른 속성의 값과 상호독립임을 가정 (Naive인 이유)
   - 독립 가정으로 인해 학습결과로 얻어진 속성(단어)의 확률값을 단어의 극성을 분류하는 용도로 사용 가능

절차 5: Polarity Lexicon 생성 - Market approach

Naive-bayes classifier(NBC)를 활용한 단어의 극성 분류 절차

1. 전체 문서를 대상으로 문서 발표 당일 콜금리가 상승할 경우 Positive(Hawkish), 하락할 경우 Negative(Dovish)로 개별 문장을 구분
2. 개별 문장의 5-gram을 생성하여 그 문장의 속성(feature)으로 사용
3. 전체 문장을 임으로 9:1의 비율로 학습과 평가용으로 나눈 후, 학습 및 평가 진행
4. 학습에서 얻어진 각 5-gram의 positive와 negative 조건부 확율의 비율로 polarity score 산출
5. 위 작업을 20번 수행하여 얻어진 polarity score의 평균을 최종 점수로 사용
6. 문장 단위 polarity score를 일별로 합산한 점수를 바탕으로 위에서 분류한 단어 리스트의 positive/negative 분류 정확도 검증

절차 5: Polarity Lexicon 생성 - Market approach

Naive-bayes classifier 관련 지표

- 분류 기준 지표
  - polarity score: 최종적으로 단어를 극성을 정하는 기준이 되는 점수 (1보다 크면 positive, 작으면 negative)

$$polarity\ score = \frac{p(feature|positive)}{p(feature|negative)}$$

  - intensity: positive와 negative의 상대강도를 측정 (큰 확률을 분자로 놓고 계산), 분류가 모호한 단어 제거를 위해 사용 (cutoff 1.3 사용)
- 평가 지표
  - precision: 분류기가 positive로 예측한 sample 중에서 맞게 예측한 것의 비율
  - recall: 전체 정보($T_p + F_n$)중에서 검출된 것(T_p)의 비율

$$precision = \frac{T_p}{T_p + F_p},\ recall = \frac{T_p}{T_p + F_n}$$

$$T_p : true\ positives, F_p : false\ positives, F_n : false\ negatives$$

절차 5: Polarity Lexicon 생성 - Market approach

분류 성능 평가 결과

- 20번의 개별 학습의 경우 평균 정확도는 약 70%
  - positive precision: 73%, positive recall 70%
  - negative precision: 68%, negative recall 72%
- 문장별 점수를 합산하여 구한 일별 점수로 평가한 polarity lexicon의 정확도는 82%
  - positive precision: 84%, positive recall 81%
  - negative precision: 79%, negative recall 83%
- 최종 n-gram수는 positive 18,685개, negative 21,280개
- 오른쪽 그림에서 5-gram feature의 높은 분류 정확도에도 불구하고 단어 단위로는 발생빈도가 크게 차이가 나지 않는 것을 알 수 있음

Figure 6: Wordcloud of word occurrences

절차 5: Polarity Lexicon 생성 - Lexical approach

1. Corpus-based approach
   - 동시발생 빈도를 사용하여 구축한 단어 그래프에서 seed word를 출발 node로 하는 label propagation 방법을 적용
   - 특정 도메인의 문서를 코퍼스로 사용하면 도메인 특화 사전을 구축하기 용이함
   - Hamilton et al. (2016)[4] 은 seed 선정에 따라 결과가 크게 달라질 수 있는 문제점을 해결하고 최신 word embedding을 적용한 SentProp framework를 제안함
   - SentProp은 단어 그래프 네트워크를 word embedding으로 부터 구하고, bootstrap-sampling approach를 사용하여 seed의 자의성과 그에 따른 결과의 불안정성 문제를 해결
2. Dictionary-based approach
   - WordNet과 같이 언어학 전문가가 수작업으로 검수한 단어 네트워크를 바탕으로 seed propagation 수행하는 방법
   - 전문가가 검수한 자료를 사용하기 때문에 일반적으로 양질의 결과를 얻을 수 있음
   - 하지만, 특정 도메인과 언어를 위한 WordNet이 존재하지 않는 경우 활용할 방법이 없음

절차 5: Polarity Lexicon 생성 - Lexical approach

Seed propagation을 활용한 단어의 극성 분류 절차

1. Word embedding 생성
   - 분류되지 않은 전체 도메인 문서(232,658개)에 대한 학습을 통해 word embedding을 통해 단어간 관계 그래프 구축
   - 기본적으로 word2vec과 같은 word embedding은 단어 단위로 구성되나, 여기서는 ngram2vec[10] 으로 n-gram에 대한 embedding을 구성
   - 5-gram center words, 5-gram context words, window size 5의 조건으로 300차원의 벡터 산출
   - 총 344,293개의 n-gram에 대해 410,902,512의 조합을 가지고 학습을 진행
2. Seed propagation 수행
   - 각각 25개로 구성된 positive와 negative seed 리스트에서 임의의 10개를 추출하여 50번의 bootstrap-sampling 실시
   - 각 단어의 polarity score는 seed에서 그 단어에 도달하는 random walk 확률을 의미
3. 단어(n-gram)의 polarity score를 산출하고 단어의 극성을 분류
   - Seed propagation을 통해 seed를 기준으로 각 단어의 positive 확률과 negative 확률이 얻어짐
   - 두 확률의 비율로 각 단어의 최종 polarity score 산출
   - positive와 negative의 상대강도를 기준으로 분류가 모호한 단어를 제거(cutoff 1.1 사용)

절차 5: Polarity Lexicon 생성 - Lexical approach

Table 2: Seed words for polarity induction

	Positive		Negative
높/VA	팽창/NNG	낮/VA	축소/NNG
인상/NNG	매파/NNG	인하/NNG	비둘기/NNG
성장/NNG	투기/NNG;억제/NNG	둔화/NNG	악화/NNG
상승/NNG	인플레이션/NNG;압력/NNG	하락/NNG	회복/NNG;못하/VX
증가/NNG	위험/NNG;선호/NNG	감소/NNG	위험/NNG;회피/NNG
상회/NNG	물가/NNG;상승/NNG	하회/NNG	물가/NNG;하락/NNG
과열/NNG	금리/NNG;상승/NNG	위축/NNG	금리/NNG;하락/NNG
확장/NNG	상방/NNG;압력/NNG	침체/NNG	하방/NNG;압력/NNG
긴축/NNG	변동성/NNG;감소/NNG	완화/NNG	변동성/NNG;확대/NNG
흑자/NNG	채권/NNG;가격/NNG;하락/NNG	적자/NNG	채권/NNG;가격/NNG;상승/NNG
견조/NNG	요금/NNG;인상/NNG	부진/NNG	요금/NNG;인하/NNG
낙관/NNG	부동산/NNG;가격/NNG;상승/NNG	비관/NNG	부동산/NNG;가격/NNG;하락/NNG
상향/NNG	(총 25개)	하향/NNG	(총 25개)

Seed propagation 결과로 얻어진 n-gram수는 positive 11,710개, negative 12,246개

절차 5: Polarity Lexicon 생성 - Supervised learning approach

Supervised learning approach

- 단어사전을 구축하는 앞의 두가지 방식과 달리 전문가에 의해 극성이 분류된 문서에 대한 학습을 통해 극성 분류기(classifier) 구축하는 방식
- 감독기반 기계학습 방식으로 비교적 적은 문서를 가지고 우수한 분류 능력을 확보할 수 있음
- 소량의 대표 샘플에 기초하기 때문에 넓은 범위의 문서 보다는 비슷한 유형의 문서를 분류하는 것에 적합

분석 대상 문서 준비

- 분석대상 문서: 금통위 직후 총재기자간담회에서 나온 총재모두발언 2,341 문장
- 문장 분류: 한국은행 내부 전문가가 모든 문장을 positive/negative/neutral로 분류(tagging)

절차 5: Polarity Lexicon 생성 - Supervised learning approach

Naive-bayes classifier 구축 절차

1. 개별 문장을 형태소 분석하고 명사, 형용사, 부사, 부정어 등을 추출하고 5-gram을 생성하여 classifier의 feature 리스트로 사용
2. feature list 중에서 극성 분류와의 연관성(bigram association with sentiment tag)이 특정 비율(best words ratio) 이상인 feature 추출
3. 전체 문장을 특정 비율(train ratio)로 random sampling하여 학습용과 평가용으로 나눈 후, 학습 및 평가 진행
4. 위 작업을 bagging(bootstrap aggregating, 30번)하여 얻어진 평가점수를 바탕으로 feature 선정기준 및 parameter 최종 선정 (best words ratio: 0.8, train ratio: 0.8)
5. 완성된 classifier를 사용하여 분석을 원하는 임의의 문서의 극성을 분류 (positive/negative 분류확률 비율이 1.5이하면 neutral로 분류)

절차 5: Polarity Lexicon 생성 - Supervised learning approach

분류 성능 평가 결과

- 30번의 개별 학습의 경우 평균 정확도는 약 86% (학습에 사용하지 않은 나머지 문장을 평가한 경우)
  - positive precision: 90%, positive recall 84%
  - negative precision: 82%, negative recall 88%
- 최종 적용한 classifier로 전체문장(2,341개)을 평가한 문장분류 정확도는 92% (in-sample)
  - positive precision: 89%, positive recall 96%
  - negative precision: 95%, negative recall 87%
- 앞서 구축한 두가지 방식의 단어사전으로 전문가가 분류한 문장(2,341개)의 극성을 평가한 분류 정확도 비교 (out-of-sample)
  - Market approach - accuracy: 68% (positive precision: 63%, negative precision: 74%)
  - Lexical approach - accuracy: 67% (positive precision: 69%, negative precision: 65%)

Monetary Sentiment Analysis

Sentiment Analysis - Measuring Minutes Sentiments

1. 개별 문장의 톤(tone) 측정
   - $Tone_{mkt}$: market approach 기반 사전을 사용하여 측정한 점수
   - $Tone_{lex}$: lexical approach 기반 사전을 사용하여 측정한 점수
   - $Tone_{nbc}$: Supervised learning approach 기반 사전을 사용하여 측정한 점수 (Naive bayes classifier 사용)
   - $Tone_{ksa}$: 서울대학교 IDS 연구실에서 만든 형태소 분석기 꼬꼬마[7]와 감성사전을 사용하여 측정한 점수
   - 각 문장의 polarity 점수는 positive와 negative n-gram의 발생빈도 차이를 이용하는 Lydia[3] 방식을 사용

$$polarity\ score\ of\ sentence = \frac{No.\ of\ positive\ n\text{-}grams - No.\ of\ negative\ n\text{-}grams}{No.\ of\ positive\ n\text{-}grams + No.\ of\ negative\ n\text{-}grams}$$

2. 의사록 센티멘트 측정
   - 의사록 전체의 센티멘트는 의사록을 구성하는 모든 문장의 센티멘트 점수를 합산하며, 산출하며 방식은 문장과 마찬가지로 Lydia 방식을 사용

$$polarity\ score\ of\ document = \frac{No.\ of\ positive\ sentences - No.\ of\ negative\ sentences}{No.\ of\ positive\ sentences + No.\ of\ negative\ sentences}$$

Economic Data Definition

분석기간: 2005. 4 - 2017. 4

- Base 용: 한국은행 기준금리
- CPI: 소비자물가지수(2015=100)(전국) 총지수
- IP: 산업생산지수(계절변동조정)
- Unemployment: 실업률(계절변동조정)
- CCI: 경기종합지수 동행지수 순환변동치
- LCI: 경기종합지수 선행지수 순환변동치
- CD Rate: CD유통수익률(91일)
- FX: 원화의 대미달러(종가) 평균자료

* 출처: 한국은행 ECOS, 통계청

Descriptive Statistics - Correlation Heatmap

Correlation among monetray sentiments and economic variables

	Base rate	CPI	IP	CD rate	FX	CCI	LCI	Unemployment	$Tone_{mkt}$	$Tone_{lex}$	$Tone_{nbc}$	$Tone_{ksa}$
Base rate												
CPI	-0.74											
IP	-0.61	0.95										
CD rate	0.98	-0.77	-0.66									
FX	-0.62	0.43	0.26	-0.50								
CCI	0.42	0.01	0.19	0.38	-0.49							
LCI	-0.12	0.04	0.17	-0.15	-0.40	0.44						
Unemployment	-0.53	0.08	-0.03	-0.49	0.26	-0.36	0.27					
$Tone_{mkt}$	0.52	-0.70	-0.54	0.50	-0.54	0.22	0.39	0.06				
$Tone_{lex}$	0.32	-0.42	-0.27	0.26	-0.59	0.26	0.49	0.07	0.85			
$Tone_{nbc}$	-0.04	-0.10	0.02	-0.07	-0.33	0.28	0.57	0.21	0.49	0.55		
$Tone_{ksa}$	-0.17	0.06	0.06	-0.25	-0.20	-0.04	0.28	0.23	0.06	0.17	0.02	

Explaining Monetary Policy Decisions - Descriptive Statistics

Table 3: Descriptive statistics of economic variables

Statistic	Mean	Median	St. Dev.	Min	Max
BOK_t	2.889	2.750	1.156	1.250	5.250
ΔBOK_t	−0.006	0.000	0.143	−1.000	0.250
π_t	0.302	0.148	1.209	−1.632	3.736
$\Delta\pi_t$	−0.011	−0.012	0.351	−0.889	0.852
$(y_t - y^*)$	−0.001	0.003	0.041	−0.228	0.065
$\Delta(y_t - y^*)$	−0.0003	0.001	0.024	−0.116	0.066
$\Delta\pi_t^e$	−0.003	0.000	0.132	−0.300	0.400
Δy_t^e	0.009	0.000	0.274	−0.600	1.000
$Tone_{mkt}$	−0.067	−0.111	0.464	−0.867	0.842
$Tone_{lex}$	−0.073	−0.099	0.293	−0.800	0.679
$Tone_{nbc}$	−0.385	−0.392	0.155	−0.806	0.053
$Tone_{ksa}$	0.549	0.554	0.101	0.178	0.818

Explaining Monetary Policy Decisions - Correlation Statistics

Table 4: Correlation matrix of economic variables

	BOK_t	ΔBOK_t	π_t	$\Delta\pi_t$	$(y_t - y^*)$	$\Delta(y_t - y^*)$	$\Delta\pi_t^e$	Δy_t^e	$Tone_{mkt}$	$Tone_{lex}$	$Tone_{nbc}$
BOK_t											
ΔBOK_t	0.19*										
π_t	0.57***	0.02									
$\Delta\pi_t$	0.07	0.10	0.16								
$(y_t - y^*)$	0.23**	0.65***	0.02	0.22*							
$\Delta(y_t - y^*)$	-0.14	0.28**	-0.09	0.06	0.28**						
$\Delta\pi_t^e$	0.18*	0.17*	0.29***	0.20*	0.23**	0.02					
Δy_t^e	-0.24**	0.13	-0.26**	-0.08	-0.02	0.32***	-0.20*				
$Tone_{mkt}$	0.55***	0.51***	0.42***	0.14	0.40***	0.07	0.12	0.03			
$Tone_{lex}$	0.39***	0.55***	0.18*	0.21*	0.52***	0.06	0.12	0.07	0.86***		
$Tone_{nbc}$	0.01	0.38***	-0.02	0.26**	0.45***	0.15	-0.04	0.21*	0.48***	0.53***	
$Tone_{ksg}$	-0.17*	0.23**	-0.39***	0.06	0.19*	0.08	0.07	0.27**	0.06	0.17*	0.03

Methodology - Taylor Rule with Ordered Probit

Taylor rule을 사용하여 통화정책 결정 설명력 분석: Ordered Probit

- $MPD_t = \alpha + \beta_i Tone_{i,t} + \rho MPD_{t-1} + \epsilon_t$
- $MPD_{t+m} = \alpha + \beta_i Tone_{i,t} + \rho MPD_t + \epsilon_t$
- $MPD_t = \alpha + \beta_i Tone_{i,t} + \gamma_1(\pi_t - \pi^*) + \gamma_2(y_t - y^*) + \gamma_3 \pi_t^e + \gamma_4 y_t^e + \rho MPD_{t-1} + \epsilon_t$
- $MPD_{t+m} = \alpha + \beta_i Tone_{i,t} + \gamma_1(\pi_t - \pi^*) + \gamma_2(y_t - y^*) + \gamma_3 \pi_t^e + \gamma_4 y_t^e + \rho MPD_t + \epsilon_t$
- MPD_t: t시점 한국은행 기준금리 변경 크기에 따라 범주화 -1:인하($<= -25bp$), 0: 변경없음, 1: 인상 ($>= +25bp$)
- $Tone_{i,t}$: 금통위 의사록 센티멘트 점수 (mkt: market approach, lex: lexical approach)
- π_t^e: 기대 인플레이션 (한은조사), y_t^e: 경기종합지수 선행지수 순환변동치
- $(\pi_t - \pi^*)$: inflation gap, π^*: 목표 물가상승률 (2%)
- $(y_t - y^*)$: output gap, y^*: hpfilter로 뽑아낸 trend

Explaining Monetary Policy Decisions - Ordered Probit Models

Table 5: Results from ordered probit models of monetary sentiments score at t_0

	Dependent variable: MPD_t						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
MPD_{t-1}	0.611**	0.639**	0.612**	−0.460	−0.349	0.363	0.624**
	(2.094)	(2.150)	(2.039)	(−1.095)	(−0.876)	(1.133)	(2.049)
$\Delta(\pi_t - \pi^*)$		0.227	0.113	−0.370	−0.433	−0.211	0.113
		(0.654)	(0.316)	(−0.688)	(−0.955)	(−0.543)	(0.316)
$\Delta(y_t - y^*)$		9.217*	8.044	7.128	10.667	7.111	7.964
		(1.923)	(1.576)	(1.158)	(1.630)	(1.327)	(1.565)
$\Delta\pi_t^e$			1.487	1.549	1.322	1.765*	1.513
			(1.544)	(1.170)	(1.137)	(1.773)	(1.561)
Δy_t^e			0.293	−0.111	−0.209	−0.001	0.331
			(0.626)	(−0.170)	(−0.376)	(−0.003)	(0.673)
$Tone_{mkt,t}$				4.805***			
				(4.473)			
$Tone_{lex,t}$					4.073***		
					(5.195)		
$Tone_{nbc,t}$						2.499***	
						(2.703)	
$Tone_{ksa,t}$							−0.313
							(−0.247)
Pseudo R^2	0.033	0.066	0.084	0.412	0.335	0.136	0.085
Observations	132	132	132	132	132	132	132

Note: * p<0.1; ** p<0.05; *** p<0.01

Explaining Monetary Policy Decisions - Ordered Probit Models

Table 6: Results from ordered probit models of monetary sentiments score at t_1

	Dependent variable: MPD_{t+1}						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
MPD_t	0.611**	0.496	0.449	−1.726***	−0.936*	0.231	0.438
	(2.094)	(1.611)	(1.416)	(−2.668)	(−1.894)	(0.687)	(1.358)
$\Delta(\pi_t - \pi^*)$		0.964***	0.995***	1.534**	0.924**	0.747*	0.967**
		(2.585)	(2.612)	(2.550)	(2.011)	(1.869)	(2.534)
$\Delta(y_t - y^*)$		6.257	3.956	10.266	7.380	3.428	4.066
		(1.272)	(0.750)	(1.407)	(1.120)	(0.627)	(0.753)
$\Delta\pi_t^e$			0.414	0.827	0.211	0.762	0.229
			(0.429)	(0.638)	(0.191)	(0.764)	(0.235)
Δy_t^e			0.654	0.927	0.443	0.451	0.458
			(1.350)	(1.368)	(0.816)	(0.896)	(0.907)
$Tone_{mkt,t}$				4.851***			
				(4.402)			
$Tone_{lex,t}$					3.861***		
					(4.986)		
$Tone_{nbc,t}$						2.042**	
						(2.243)	
$Tone_{ksa,t}$							1.868
							(1.432)
Pseudo R^2	0.033	0.097	0.109	0.414	0.319	0.143	0.123
Observations	132	132	132	132	132	132	132

Note: * p<0.1; ** p<0.05; *** p<0.01

Explaining Monetary Policy Decisions - Ordered Probit Models

Table 7: Results from ordered probit models of monetary sentiments score at t_2

	Dependent variable: MPD_{t+2}						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
MPD_t	0.706**	0.620**	0.606**	−0.291	−0.071	0.340	0.611**
	(2.503)	(2.132)	(2.061)	(−0.772)	(−0.199)	(1.093)	(2.079)
$\Delta(\pi_t - \pi^*)$		0.459	0.444	0.327	0.219	0.043	0.459
		(1.304)	(1.242)	(0.829)	(0.572)	(0.109)	(1.278)
$\Delta(y_t - y^*)$		6.000	5.605	7.133	6.834	4.013	5.781
		(1.195)	(1.054)	(1.261)	(1.206)	(0.699)	(1.090)
$\Delta\pi_t^e$			0.284	0.495	0.277	0.767	0.339
			(0.308)	(0.499)	(0.287)	(0.788)	(0.366)
Δy_t^e			0.116	0.159	0.002	−0.214	0.224
			(0.246)	(0.313)	(0.005)	(−0.427)	(0.455)
$Tone_{mkt,t}$				1.645***			
				(3.793)			
$Tone_{lex,t}$					1.997***		
					(3.467)		
$Tone_{nbc,t}$						3.900***	
						(3.935)	
$Tone_{ksa,t}$							−0.953
							(−0.764)
Pseudo R^2	0.047	0.072	0.073	0.197	0.162	0.189	0.077
Observations	131	131	131	131	131	131	131

Note: * p<0.1; ** p<0.05; *** p<0.01

Methodology - Taylor Rule with OLS

Taylor rule을 사용하여 통화정책 결정 설명력 분석: OLS

- $BOK_t = \alpha + \beta_i Tone_{i,t} + \rho BOK_{t-1} + \epsilon_t$
- $BOK_{t+m} = \alpha + \beta_i Tone_{i,t} + \rho BOK_t + \epsilon_t$
- $BOK_t = \alpha + \beta_i Tone_{i,t} + \gamma_1(\pi_t - \pi^*) + \gamma_2(y_t - y^*) + \gamma_3 \pi_t^e + \gamma_4 y_t^e + \rho BOK_{t-1} + \epsilon_t$
- $BOK_{t+m} = \alpha + \beta_i Tone_{i,t} + \gamma_1(\pi_t - \pi^*) + \gamma_2(y_t - y^*) + \gamma_3 \pi_t^e + \gamma_4 y_t^e + \rho BOK_t + \epsilon_t$
- BOK_t: t시점 한국은행 기준금리
- $Tone_{i,t}$: 금통위 의사록 센티멘트 점수 (mkt: market approach, lex: lexical approach)
- π_t^e: 기대 인플레이션 (한은조사), y_t^e: 경기종합지수 선행지수 순환변동치
- $(\pi_t - \pi^*)$: inflation gap, π^*: 목표 물가상승률 (2%)
- $(y_t - y^*)$: output gap, y^*: hpfilter로 뽑아낸 trend

Explaining Monetary Policy Decisions - OLS

Table 8: Results from OLS of monetary sentiments score at t_0

	Dependent variable: ΔBOK_t						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ΔBOK_{t-1}	0.394***	0.427***	0.415***	0.257***	0.217***	0.337***	0.394***
	(4.893)	(5.570)	(5.426)	(3.296)	(2.767)	(4.240)	(5.066)
$\Delta(\pi_t - \pi^*)$		0.019	0.011	−0.003	−0.017	−0.014	0.009
		(0.628)	(0.342)	(−0.114)	(−0.597)	(−0.450)	(0.296)
$\Delta(y_t - y^*)$		1.929***	1.790***	1.559***	1.591***	1.619***	1.776***
		(4.204)	(3.689)	(3.456)	(3.604)	(3.397)	(3.670)
$\Delta\pi_t^e$			0.155*	0.124	0.123	0.177**	0.141
			(1.824)	(1.567)	(1.593)	(2.129)	(1.644)
Δy_t^e			0.033	0.030	0.017	0.012	0.017
			(0.770)	(0.760)	(0.427)	(0.275)	(0.376)
$Tone_{mkt,t}$				0.114***			
				(4.737)			
$Tone_{lex,t}$					0.207***		
					(5.319)		
$Tone_{nbc,t}$						0.215***	
						(2.788)	
$Tone_{ksa,t}$							0.155
							(1.317)
Observations	132	132	132	132	132	132	132
Adjusted R^2	0.149	0.245	0.254	0.362	0.387	0.292	0.258

Note: * $p<0.1$; ** $p<0.05$; *** $p<0.01$

Explaining Monetary Policy Decisions - OLS

Table 9: Results from OLS of monetary sentiments score at t_1

	Dependent variable: ΔBOK_{t+1}						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ΔBOK_t	0.007	0.012	0.015	−0.026**	−0.013	0.012	0.018*
	(0.655)	(1.195)	(1.417)	(−2.337)	(−1.295)	(1.214)	(1.753)
$\Delta(\pi_t - \pi^*)$		0.080**	0.080**	0.057*	0.043	0.046	0.075**
		(2.450)	(2.427)	(1.968)	(1.483)	(1.400)	(2.326)
$\Delta(y_t - y^*)$		2.194***	1.843***	1.525***	1.670***	1.693***	1.884***
		(4.530)	(3.656)	(3.448)	(3.839)	(3.510)	(3.810)
$\Delta\pi_t^e$			0.093	0.079	0.066	0.115	0.061
			(1.039)	(1.005)	(0.852)	(1.341)	(0.685)
Δy_t^e			0.097**	0.055	0.050	0.065	0.067
			(2.144)	(1.377)	(1.269)	(1.489)	(1.468)
$Tone_{mkt,t}$				0.168***			
				(6.380)			
$Tone_{lex,t}$					0.251***		
					(6.697)		
$Tone_{nbc,t}$						0.274***	
						(3.683)	
$Tone_{ksa,t}$							0.289**
							(2.474)
Observations	132	132	132	132	132	132	132
Adjusted R^2	−0.004	0.164	0.184	0.379	0.394	0.258	0.216

Note: * p<0.1; ** p<0.05; *** p<0.01

Explaining Monetary Policy Decisions - OLS

Table 10: Results from OLS of monetary sentiments score at t_2

	Dependent variable: ΔBOK_{t+2}						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ΔBOK_t	−0.005	−0.001	0.0005	−0.038***	−0.022*	−0.004	0.001
	(−0.466)	(−0.107)	(0.041)	(−3.077)	(−1.941)	(−0.363)	(0.124)
$\Delta(\pi_t - \pi^*)$		0.064*	0.068*	0.045	0.037	0.020	0.066*
		(1.845)	(1.910)	(1.389)	(1.103)	(0.583)	(1.864)
$\Delta(y_t - y^*)$		1.581***	1.525***	1.194**	1.363***	1.294**	1.537***
		(3.082)	(2.799)	(2.403)	(2.697)	(2.570)	(2.812)
$\Delta\pi_t^e$			−0.040	−0.051	−0.060	−0.008	−0.049
			(−0.414)	(−0.590)	(−0.678)	(−0.093)	(−0.504)
Δy_t^e			0.019	−0.019	−0.018	−0.023	0.010
			(0.385)	(−0.426)	(−0.394)	(−0.513)	(0.201)
$Tone_{mkt,t}$				0.158***			
				(5.357)			
$Tone_{lex,t}$					0.204***		
					(4.688)		
$Tone_{nbc,t}$						0.378***	
						(4.879)	
$Tone_{ksa,t}$							0.084
							(0.654)
Observations	131	131	131	131	131	131	131
Adjusted R^2	−0.006	0.077	0.066	0.235	0.200	0.210	0.061

Note: * p<0.1; ** p<0.05; *** p<0.01

Comparison with English Sentiment Aanalysis

Sentiment Analysis - English Translation Approach

1. 영문 변역
   - 영어의 경우 금융을 포함한 다양한 분야의 Sentiment 측정 방법이 다수 존재
   - Google Cloud Translation API을 사용하여 의사록 모든 문장을 영문으로 번역
2. 각 문장의 톤(tone) 측정
   - $Tone_{LM}$: Loughran and McDonald (2011) 사전
   - $Tone_{HIV4}$: Harvard IV-4 사전
   - $Tone_{Google}$: Google Cloud Natural Language API의 Sentiment Analysis를 사용
3. 의사록 센티멘트 측정
   - 의사록을 구성하는 모든 문장의 센티멘트 점수를 합산하여 산출

$$polarity\ score = \frac{No.\ of\ positive\ sentences - No.\ of\ negative\ sentences}{No.\ of\ positive\ sentences + No.\ of\ negative\ sentences}$$

Descriptive Statistics - Correlation Heatmap

Correlation among monetray sentiments and economic variables

	Base rate	CPI	IP	FX	CCI	LCI	$Tone_{mkt}$	$Tone_{lex}$	$Tone_{ksa}$	$Tone_{Google}$	$Tone_{HiV4}$	$Tone_{LM}$
Base rate												
CPI	-0.74											
IP	-0.61	0.95										
FX	-0.62	0.42	0.25									
CCI	0.44	-0.02	0.17	-0.51								
LCI	-0.11	0.02	0.17	-0.41	0.43							
$Tone_{mkt}$	0.52	-0.70	-0.54	-0.54	0.24	0.40						
$Tone_{lex}$	0.33	-0.42	-0.26	-0.59	0.28	0.50	0.85					
$Tone_{ksa}$	-0.17	0.03	0.04	-0.21	-0.06	0.28	0.08	0.19				
$Tone_{Google}$	-0.08	0.05	0.16	-0.27	0.09	0.40	0.26	0.37	0.20			
$Tone_{HiV4}$	-0.23	-0.03	-0.05	-0.18	-0.22	0.20	0.10	0.19	0.34	0.39		
$Tone_{LM}$	-0.10	-0.11	-0.06	-0.34	-0.03	0.42	0.38	0.53	0.41	0.62	0.57	

Explaining Monetary Policy Decisions - Descriptive Statistics

Table 11: Descriptive statistics of economic variables

Statistic	Mean	Median	St. Dev.	Min	Max
BOK_t	2.889	2.750	1.156	1.250	5.250
ΔBOK_t	−0.006	0.000	0.143	−1.000	0.250
π_t	0.302	0.148	1.209	−1.632	3.736
$\Delta\pi_t$	−0.011	−0.012	0.351	−0.889	0.852
$(y_t - y^*)$	−0.001	0.003	0.041	−0.228	0.065
$\Delta(y_t - y^*)$	−0.0003	0.001	0.024	−0.116	0.066
$\Delta\pi_t^e$	−0.003	0.000	0.132	−0.300	0.400
Δy_t^e	0.009	0.000	0.274	−0.600	1.000
$Tone_{mkt}$	−0.067	−0.111	0.464	−0.867	0.842
$Tone_{lex}$	−0.073	−0.099	0.293	−0.800	0.679
$Tone_{HIV4}$	0.420	0.456	0.142	0.030	0.733
$Tone_{LM}$	−0.397	−0.395	0.179	−0.833	0.091

Explaining Monetary Policy Decisions - Correlation Statistics

Table 12: Correlation matrix of economic variables

	BOK_t	ΔBOK_t	π_t	$\Delta\pi_t$	$(y_t - y^*)$	$\Delta(y_t - y^*)$	$\Delta\pi_t^e$	Δy_t^e	$Tone_{mkt}$	$Tone_{lex}$	$Tone_{HIV4}$
BOK_t											
ΔBOK_t	0.19*										
π_t	0.57***	0.02									
$\Delta\pi_t$	0.07	0.10	0.16								
$(y_t - y^*)$	0.23**	0.65***	0.02	0.22*							
$\Delta(y_t - y^*)$	-0.14	0.28**	-0.09	0.06	0.28**						
$\Delta\pi_t^e$	0.18*	0.17*	0.29***	0.20*	0.23**	0.02					
Δy_t^e	-0.24**	0.13	-0.26**	-0.08	-0.02	0.32***	-0.20*				
$Tone_{mkt}$	0.55***	0.51***	0.42***	0.14	0.40***	0.07	0.12	0.03			
$Tone_{lex}$	0.39***	0.55***	0.18*	0.21*	0.52***	0.06	0.12	0.07	0.86***		
$Tone_{HIV4}$	-0.22*	0.02	-0.39***	0.15	-0.02	0.08	-0.27**	0.31***	0.06	0.14	
$Tone_{LM}$	-0.07	0.26**	-0.33***	0.09	0.19*	0.07	-0.16	0.40***	0.38***	0.52***	0.54***

Explaining Monetary Policy Decisions - Ordered Probit Models

Table 13: Results from ordered probit models of monetary sentiments score at t_0

	Dependent variable: MPD_t						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
MPD_{t-1}	0.608**	0.636**	0.609**	−0.460	−0.348	0.612**	0.471
	(2.086)	(2.143)	(2.031)	(−1.094)	(−0.876)	(2.045)	(1.502)
$\Delta(\pi_t - \pi^*)$		0.226	0.113	−0.370	−0.429	0.196	0.013
		(0.653)	(0.316)	(−0.689)	(−0.948)	(0.533)	(0.036)
$\Delta(y_t - y^*)$		9.191*	8.006	7.139	10.628	7.830	8.876*
		(1.919)	(1.570)	(1.160)	(1.626)	(1.542)	(1.695)
$\Delta\pi_t^e$			1.481	1.549	1.311	1.230	1.656*
			(1.537)	(1.170)	(1.128)	(1.233)	(1.697)
Δy_t^e			0.293	−0.111	−0.204	0.445	−0.093
			(0.625)	(−0.171)	(−0.368)	(0.899)	(−0.178)
$Tone_{mkt,t}$				4.802***			
				(4.463)			
$Tone_{lex,t}$					4.060***		
					(5.180)		
$Tone_{HIV4,t}$						−0.960	
						(−0.997)	
$Tone_{LM,t}$							1.393*
							(1.692)
Pseudo R^2	0.033	0.066	0.084	0.413	0.336	0.091	0.105
Observations	131	131	131	131	131	131	131

Note: * p<0.1; ** p<0.05; *** p<0.01

Explaining Monetary Policy Decisions - Ordered Probit Models

Table 14: Results from ordered probit models of monetary sentiments score at t_1

	Dependent variable: MPD_{t+1}						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
MPD_t	0.611**	0.496	0.449	−1.726***	−0.936*	0.486	0.271
	(2.094)	(1.611)	(1.416)	(−2.668)	(−1.894)	(1.513)	(0.817)
$\Delta(\pi_t - \pi^*)$		0.964***	0.995***	1.534**	0.924**	0.886**	0.891**
		(2.585)	(2.612)	(2.550)	(2.011)	(2.278)	(2.266)
$\Delta(y_t - y^*)$		6.257	3.956	10.266	7.380	3.722	5.375
		(1.272)	(0.750)	(1.407)	(1.120)	(0.697)	(0.961)
$\Delta\pi_t^e$			0.414	0.827	0.211	0.802	0.891
			(0.429)	(0.638)	(0.191)	(0.793)	(0.891)
Δy_t^e			0.654	0.927	0.443	0.487	0.048
			(1.350)	(1.368)	(0.816)	(0.968)	(0.089)
$Tone_{mkt,t}$				4.851***			
				(4.402)			
$Tone_{lex,t}$					3.861***		
					(4.986)		
$Tone_{HIV4,t}$						1.293	
						(1.340)	
$Tone_{LM,t}$							2.440***
							(2.844)
Pseudo R^2	0.033	0.097	0.109	0.414	0.319	0.121	0.165
Observations	132	132	132	132	132	132	132

Note: * p<0.1; ** p<0.05; *** p<0.01

Explaining Monetary Policy Decisions - Ordered Probit Models

Table 15: Results from ordered probit models of monetary sentiments score at t_2

	Dependent variable: MPD_{t+2}						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
MPD_t	0.706**	0.620**	0.606**	−0.291	−0.071	0.691**	0.502*
	(2.503)	(2.132)	(2.061)	(−0.772)	(−0.199)	(2.280)	(1.658)
$\Delta(\pi_t - \pi^*)$		0.459	0.444	0.327	0.219	0.258	0.294
		(1.304)	(1.242)	(0.829)	(0.572)	(0.695)	(0.793)
$\Delta(y_t - y^*)$		6.000	5.605	7.133	6.834	5.534	6.964
		(1.195)	(1.054)	(1.261)	(1.206)	(1.015)	(1.246)
$\Delta\pi_t^e$			0.284	0.495	0.277	0.939	0.784
			(0.308)	(0.499)	(0.287)	(0.957)	(0.813)
Δy_t^e			0.116	0.159	0.002	−0.186	−0.464
			(0.246)	(0.313)	(0.005)	(−0.374)	(−0.879)
$Tone_{mkt,t}$				1.645***			
				(3.793)			
$Tone_{lex,t}$					1.997***		
					(3.467)		
$Tone_{HIV4,t}$						2.149**	
						(2.130)	
$Tone_{LM,t}$							2.272***
							(2.745)
Pseudo R^2	0.047	0.072	0.073	0.197	0.162	0.106	0.127
Observations	131	131	131	131	131	131	131

Note: * p<0.1; ** p<0.05; *** p<0.01

Explaining Monetary Policy Decisions - OLS

Table 16: Results from OLS of monetary sentiments score at t_0

	Dependent variable: ΔBOK_t						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ΔBOK_{t-1}	0.394***	0.427***	0.415***	0.257***	0.217***	0.415***	0.367***
	(4.874)	(5.548)	(5.404)	(3.282)	(2.755)	(5.380)	(4.579)
$\Delta(\pi_t - \pi^*)$		0.019	0.011	−0.003	−0.017	0.011	0.003
		(0.626)	(0.338)	(−0.117)	(−0.590)	(0.333)	(0.087)
$\Delta(y_t - y^*)$		1.930***	1.790***	1.560***	1.591***	1.790***	1.805***
		(4.187)	(3.675)	(3.443)	(3.588)	(3.659)	(3.740)
$\Delta\pi_t^e$			0.155*	0.124	0.123	0.155*	0.176**
			(1.817)	(1.564)	(1.577)	(1.740)	(2.057)
Δy_t^e			0.033	0.030	0.017	0.033	0.00004
			(0.763)	(0.750)	(0.430)	(0.735)	(0.001)
$Tone_{mkt,t}$				0.114***			
				(4.719)			
$Tone_{lex,t}$					0.207***		
					(5.299)		
$Tone_{HIV4,t}$						−0.002	
						(−0.019)	
$Tone_{LM,t}$							0.134*
							(1.859)
Observations	131	131	131	131	131	131	131
Adjusted R^2	0.149	0.245	0.253	0.362	0.386	0.247	0.268

Note: * $p<0.1$; ** $p<0.05$; *** $p<0.01$

Explaining Monetary Policy Decisions - OLS

Table 17: Results from OLS of monetary sentiments score at t_1

	Dependent variable: ΔBOK_{t+1}						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ΔBOK_t	0.007	0.012	0.015	−0.026**	−0.013	0.017	0.013
	(0.655)	(1.195)	(1.417)	(−2.337)	(−1.295)	(1.611)	(1.280)
$\Delta(\pi_t - \pi^*)$		0.080**	0.080**	0.057*	0.043	0.068**	0.063*
		(2.450)	(2.427)	(1.968)	(1.483)	(2.021)	(1.969)
$\Delta(y_t - y^*)$		2.194***	1.843***	1.525***	1.670***	1.869***	1.945***
		(4.530)	(3.656)	(3.448)	(3.839)	(3.722)	(4.030)
$\Delta\pi_t^e$			0.093	0.079	0.066	0.126	0.127
			(1.039)	(1.005)	(0.852)	(1.369)	(1.480)
Δy_t^e			0.097**	0.055	0.050	0.079*	0.032
			(2.144)	(1.377)	(1.269)	(1.701)	(0.681)
$Tone_{mkt,t}$				0.168***			
				(6.380)			
$Tone_{lex,t}$					0.251***		
					(6.697)		
$Tone_{HIV4,t}$						0.129	
						(1.454)	
$Tone_{LM,t}$							0.247***
							(3.606)
Observations	132	132	132	132	132	132	132
Adjusted R^2	−0.004	0.164	0.184	0.379	0.394	0.191	0.255

Note: * p<0.1; ** p<0.05; *** p<0.01

Explaining Monetary Policy Decisions - OLS

Table 18: Results from OLS of monetary sentiments score at t_2

	Dependent variable: ΔBOK_{t+2}						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ΔBOK_t	−0.005	−0.001	0.0005	−0.038***	−0.022*	0.004	−0.002
	(−0.466)	(−0.107)	(0.041)	(−3.077)	(−1.941)	(0.392)	(−0.221)
$\Delta(\pi_t - \pi^*)$		0.064*	0.068*	0.045	0.037	0.045	0.048
		(1.845)	(1.910)	(1.389)	(1.103)	(1.262)	(1.395)
$\Delta(y_t - y^*)$		1.581***	1.525***	1.194**	1.363***	1.569***	1.608***
		(3.082)	(2.799)	(2.403)	(2.697)	(2.945)	(3.089)
$\Delta\pi_t^e$			−0.040	−0.051	−0.060	0.023	−0.0003
			(−0.414)	(−0.590)	(−0.678)	(0.242)	(−0.003)
Δy_t^e			0.019	−0.019	−0.018	−0.015	−0.051
			(0.385)	(−0.426)	(−0.394)	(−0.299)	(−1.015)
$Tone_{mkt,t}$				0.158***			
				(5.357)			
$Tone_{lex,t}$					0.204***		
					(4.688)		
$Tone_{HIV4,t}$						0.246***	
						(2.640)	
$Tone_{LM,t}$							0.270***
							(3.648)
Observations	131	131	131	131	131	131	131
Adjusted R^2	−0.006	0.077	0.066	0.235	0.200	0.108	0.149

Note: * p<0.1; ** p<0.05; *** p<0.01

Topic Sentiment Analysis

Topic Modeling - Methodology

- Latent Variables
  - The implicit assumption of the dictionary $\mathfrak{D} = \{labor, wage, employ\}$ is that each word maps back into an underlying topic "labor markets"
  - We cannot observe the topics in text, only observe the words that those topics tend to generate
  - A natural way forward is to model topics with latent variables
- Latent variable models generally share the following features:
  - Associate each word in the vocabulary to any given latent variable
  - Allow each word to have associations with multiple topics
  - Associate each document with topics
- Topic modeling methodologies
  - LSA: Latent Semantic Analysis - SVD
  - LDA: Latent Dirichlet Allocation

Topic Modeling - Latent Dirichlet Allocation

- Latent Dirichlet Allocation (LDA) model
  - Blei, Ng and Jordan (2003)[1] cited 23,300+ times
  - LDA (and its extensions) estimates what fraction of each document in a collection is devoted to each of several "topics"
  - LDA is an unsupervised learning approach - we don't set probabilities
- Modeling procedure
  1. Prepare documents: number of documents M, number of words in document d N_d
  2. Tell the model how many topics (K) there should be
     - Perplexity scores
  3. Model will generate φ_k **topic distributions**: $\varphi_k \sim Dirichlet_V(\beta)$
     - the distribution over words for each topic
  4. Model also generates θ_d **document distributions**: $\theta_d \sim Dirichlet_K(\alpha)$

Topic Modeling - 분석 절차

1. 통화정책 관련 텍스트 준비
   - 앞서 사용한 뉴스, 애널리스트 분석리포트, 금통위 의사록 등 통화정책 관련 텍스트 232,658건 사용
   - 형태소 분석 후, 명사만 추출
2. LDA 모형 추정
   - 모든 문서의 각 문장을 Bag-of-Words의 벡터 형태로 변환
   - 모형 추정 (적용 parameter: decay=0.5, offset=64)[5], 10 passes, 1000 iteration for E-step
   - 토픽 수를 10에서 40까지 2씩 증가시키며 모형 추정 반복
3. 최적 토픽 수 결정 및 토픽명 부여
   - Coherence measure 와 wordcloud, dendrogram 등을 참고하여 최적 토픽 수 결정
   - 각 토픽의 이름 부여
4. 토픽 분포 (φ_k) 분석
   - 전체 문서의 토픽 분포 정보를 집계하여 토픽 비중 산출

Topic Modeling Results - Number of Topics (K)

- Topic coherence measure와 wordcloud, dendrogram을 검토한 결과, 토픽의 수를 36개로 결정함

Topic Modeling Results - Topic Wordcloud (φ_k)

Topic Modeling Results - Topic Clustering (φ_k)

Topic Modeling Results - Document Distribution (θ_d)

Table 19: 평균 토픽 비중

No.	Topic name	Total	Minutes	News	Report	No.	Topic name	Total	Minutes	News	Report
1	환율	5.27	11.20	5.94	3.77	19	금융상품	2.36	0.22	3.42	0.75
2	금융산업정책	2.70	2.24	3.15	1.99	20	기업가치평가	1.65	0.27	2.13	0.96
3	채권발행시장1	3.28	0.73	1.29	6.76	21	자본규제	0.77	0.42	0.48	1.26
4	통화정책	3.83	12.56	4.47	2.19	22	국채선물시황	4.69	0.46	4.44	5.38
5	채권발행시장2	2.79	1.32	2.67	3.09	23	국내증시	0.82	0.17	1.21	0.22
6	신용위기	1.80	1.03	2.09	1.38	24	중소기업자금	1.35	0.65	2.06	0.22
7	채권스왑시황	4.30	3.05	4.02	4.85	25	채권시세	0.99	0.33	0.52	1.81
8	물가전망	3.27	10.56	2.68	3.79	26	캐리트레이드	1.42	0.66	1.87	0.71
9	신용등급	1.41	0.38	0.93	2.26	27	정부규제	1.44	0.87	1.75	0.96
10	부동산투자	1.21	0.15	1.71	0.46	28	펀드투자자금	4.56	2.19	5.40	3.32
11	글로벌국채시황	2.17	0.40	1.34	3.66	29	기업구조조정	1.43	1.66	1.59	1.14
12	대외건전성	1.26	2.74	1.16	1.33	30	가계소비/소득	1.93	5.68	1.72	2.03
13	부동산정책	3.03	2.44	3.99	1.48	31	유동성공급	0.61	0.21	0.46	0.90
14	유로존	3.97	1.27	3.33	5.20	32	정부/정치	2.08	0.65	1.51	3.14
15	경제성장	5.01	13.60	4.58	5.18	33	원자재가격	1.61	0.90	1.54	1.76
16	단기자금	0.89	0.79	1.09	0.56	34	채권투자전략	1.79	0.13	1.44	2.48
17	해외증시	2.12	0.37	2.74	1.21	35	가계부채	6.05	8.03	7.81	2.99
18	글로벌통화정책	5.48	4.87	4.36	7.38	36	기업신용평가	3.56	0.46	1.86	6.59

Topic Modeling Results - Document Distribution (θ_d)

Figure 7: 의사록 토픽 비중 추이

Figure 8: 섹션별 의사록 토픽 비중

Topic Modeling Results - Document Distribution (θ_d)

Figure 9: 뉴스 토픽 비중 추이

Figure 10: 리포트 토픽 비중 추이

Topic Sentiment Analysis - 분석 절차

1. 분석대상 코퍼스 선정
   - 분석을 원하는 주제와 관련된 문서를 준비하고 전처리 작업을 통해 텍스트 추출
   - 여기서는 금통위 의사록이 분석 대상
2. 문장별 센티멘트 점수 산출
   - 모든 문장의 센티멘트 점수 산출
   - ekonlpy.sentiment의 MPKO 클래스 사용
3. 문장별 토픽 비중 산출
   - 문장별 토픽 분포 산출
   - ekonlpy.topic의 MPTK 클래스 사용
4. 토픽 재분류
   - 총 36개의 토픽 중에서 유의미한 토픽 선정 후, 목적에 적합하게 통합 재분류
   - 금통위 의사록의 경우, 4가지 토픽으로 재분류 (경제현황, 대외여건, 통화정책방향, 글로벌통화정책)
5. 토픽 센티멘트 점수 산출
   - 각 문장의 센티멘트 점수를 그 문장의 토픽 비중으로 가중평균하여 토픽 센티멘트 점수 산출

Topic Sentiment Analysis - 토픽 재분류

Table 20: 토픽 재분류

No.	Topic name	Symbol	Name
0	환율	FC	대외여건
3	통화정책	FG	통화정책방향
5	신용위기	FC	대외여건
7	물가전망	FG	통화정책방향
11	대외건전성	FC	대외여건
12	부동산정책	EC	경제현황
13	유로존	FC	대외여건
14	경제성장	EC	경제현황
17	글로벌통화정책	FC	대외여건
29	가계소비/소득	EC	경제현황
30	유동성공급	FC	대외여건
31	정부/정치	FC	대외여건
32	원자재가격	FC	대외여건
34	가계부채	EC	경제현황

Figure 11: 섹션별 의사록 토픽 재분류 비중

Topic Sentiment Analysis - Descriptive Statistics

Table 21: Descriptive statistics of variables

Statistic	Mean	Median	St. Dev.	Min	Max
BOK_t	2.889	2.750	1.156	1.250	5.250
ΔBOK_t	−0.006	0.000	0.143	−1.000	0.250
π_t	0.302	0.148	1.209	−1.632	3.736
$\Delta\pi_t$	−0.011	−0.012	0.351	−0.889	0.852
y_t	−0.001	0.003	0.041	−0.228	0.065
Δy_t	−0.0003	0.001	0.024	−0.116	0.066
$\Delta\pi_t^e$	−0.003	0.000	0.132	−0.300	0.400
Δy_t^e	0.009	0.000	0.274	−0.600	1.000
$Tone_{lex}$	−0.073	−0.099	0.293	−0.800	0.679
경제현황	−0.055	−0.048	0.111	−0.396	0.255
대외여건	−0.048	−0.050	0.063	−0.175	0.193
통화정책방향	0.079	0.073	0.131	−0.317	0.361

Topic Sentiment Analysis - Correlation Statistics

Table 22: Correlation of economic variables and topic sentiment scores

	BOK_t	ΔBOK_t	π_t	$\Delta\pi_t$	y_t	Δy_t	$\Delta\pi_t^e$	Δy_t^e	$Tone_{lex}$	경제현황	대외여건
BOK_t											
ΔBOK_t	0.19*										
π_t	0.57***	0.02									
$\Delta\pi_t$	0.07	0.10	0.16								
y_t	0.23**	0.65***	0.02	0.22*							
Δy_t	-0.14	0.28**	-0.09	0.06	0.28**						
$\Delta\pi_t^e$	0.18*	0.17*	0.29***	0.20*	0.23**	0.02					
Δy_t^e	-0.24**	0.13	-0.26**	-0.08	-0.02	0.32***	-0.20*				
$Tone_{lex}$	0.39***	0.55***	0.18*	0.21*	0.52***	0.06	0.12	0.07			
경제현황	0.16	0.48***	-0.14	0.25**	0.50***	0.10	0.09	0.16	0.78***		
대외여건	0.08	0.18*	0.00	0.13	0.13	0.01	0.01	0.13	0.69***	0.39***	
통화정책방향	0.52***	0.59***	0.28**	0.11	0.58***	0.01	0.16	-0.10	0.88***	0.60***	0.43***

Topic Sentiment Analysis - Ordered Probit Models

Table 23: Results from ordered probit models of monetary sentiments score at t_0

	Dependent variable: MPD_t						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
MPD_{t-1}	0.611**	0.639**	0.612**	−0.349	0.069	0.531*	−0.487
	(2.094)	(2.150)	(2.039)	(−0.876)	(0.197)	(1.734)	(−1.103)
$\Delta(\pi_t - \pi^*)$		0.227	0.113	−0.433	−0.321	−0.002	−0.160
		(0.654)	(0.316)	(−0.955)	(−0.806)	(−0.006)	(−0.341)
$\Delta(y_t - y^*)$		9.217*	8.044	10.667	9.167	8.911*	14.671*
		(1.923)	(1.576)	(1.630)	(1.589)	(1.723)	(1.860)
$\Delta\pi_t^e$			1.487	1.322	1.358	1.517	1.281
			(1.544)	(1.137)	(1.321)	(1.541)	(0.990)
Δy_t^e			0.293	−0.209	−0.319	0.123	0.903
			(0.626)	(−0.376)	(−0.612)	(0.254)	(1.473)
$Tone_{lex}$				4.073***			
				(5.195)			
경제현황					6.187***		
					(4.064)		
대외여건						4.535**	
						(2.194)	
통화정책방향							10.945***
							(5.361)
Pseudo R^2	0.033	0.066	0.084	0.335	0.211	0.118	0.385
Observations	132	132	132	132	132	132	132

Note: * p<0.1; ** p<0.05; *** p<0.01

Topic Sentiment Analysis - Ordered Probit Models

Table 24: Results from ordered probit models of monetary sentiments score at t_1

	Dependent variable: MPD_{t+1}						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
MPD_t	0.611**	0.496	0.449	−0.936*	−0.101	0.306	−1.223**
	(2.094)	(1.611)	(1.416)	(−1.894)	(−0.269)	(0.930)	(−2.352)
$\Delta(\pi_t - \pi^*)$		0.964***	0.995***	0.924**	0.726*	0.904**	1.398***
		(2.585)	(2.612)	(2.011)	(1.799)	(2.333)	(2.753)
$\Delta(y_t - y^*)$		6.257	3.956	7.380	4.945	4.889	8.660
		(1.272)	(0.750)	(1.120)	(0.859)	(0.912)	(1.186)
$\Delta\pi_t^e$			0.414	0.211	0.477	0.461	−0.147
			(0.429)	(0.191)	(0.475)	(0.465)	(−0.129)
Δy_t^e			0.654	0.443	0.350	0.469	1.377**
			(1.350)	(0.816)	(0.685)	(0.936)	(2.378)
$Tone_{lex}$				3.861***			
				(4.986)			
경제현황					4.893***		
					(3.375)		
대외여건						5.458***	
						(2.591)	
통화정책방향							9.638***
							(5.006)
Pseudo R^2	0.033	0.097	0.109	0.319	0.191	0.156	0.336
Observations	132	132	132	132	132	132	132

Note: * p<0.1; ** p<0.05; *** p<0.01

Topic Sentiment Analysis - Ordered Probit Models

Table 25: Results from ordered probit models of monetary sentiments score at t_2

	Dependent variable: MPD_{t+2}						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
MPD_t	0.706**	0.620**	0.606**	−0.071	0.249	0.523*	−0.111
	(2.503)	(2.132)	(2.061)	(−0.199)	(0.762)	(1.740)	(−0.299)
$\Delta(\pi_t - \pi^*)$		0.459	0.444	0.219	0.210	0.371	0.384
		(1.304)	(1.242)	(0.572)	(0.554)	(1.023)	(1.028)
$\Delta(y_t - y^*)$		6.000	5.605	6.834	6.120	6.263	6.770
		(1.195)	(1.054)	(1.206)	(1.091)	(1.168)	(1.184)
$\Delta\pi_t^e$			0.284	0.277	0.314	0.313	0.175
			(0.308)	(0.287)	(0.331)	(0.337)	(0.184)
Δy_t^e			0.116	0.002	−0.150	0.025	0.368
			(0.246)	(0.005)	(−0.306)	(0.051)	(0.753)
$Tone_{lex}$				1.997***			
				(3.467)			
경제현황					3.803***		
					(2.756)		
대외여건						3.050	
						(1.487)	
통화정책방향							4.211***
							(3.260)
Pseudo R^2	0.047	0.072	0.073	0.162	0.129	0.088	0.151
Observations	131	131	131	131	131	131	131

Note: * p<0.1; ** p<0.05; *** p<0.01

Topic Sentiment Analysis - OLS

Table 26: Results from OLS of monetary sentiments score at t_0

	Dependent variable: ΔBOK_t						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ΔBOK_{t-1}	0.394***	0.427***	0.415***	0.217***	0.266***	0.399***	0.176**
	(4.893)	(5.570)	(5.426)	(2.767)	(3.303)	(5.164)	(2.307)
$\Delta(\pi_t - \pi^*)$		0.019	0.011	−0.017	−0.021	0.006	0.001
		(0.628)	(0.342)	(−0.597)	(−0.688)	(0.176)	(0.022)
$\Delta(y_t - y^*)$		1.929***	1.790***	1.591***	1.647***	1.806***	1.512***
		(4.204)	(3.689)	(3.604)	(3.588)	(3.728)	(3.563)
$\Delta\pi_t^e$			0.155*	0.123	0.143*	0.155*	0.105
			(1.824)	(1.593)	(1.782)	(1.821)	(1.408)
Δy_t^e			0.033	0.017	0.003	0.025	0.059
			(0.770)	(0.427)	(0.080)	(0.592)	(1.585)
$Tone_{lex}$				0.207***			
				(5.319)			
경제현황					0.447***		
					(4.092)		
대외여건						0.219	
						(1.247)	
통화정책방향							0.539***
							(6.447)
Observations	132	132	132	132	132	132	132
Adjusted R^2	0.149	0.245	0.254	0.387	0.337	0.257	0.435

Note: * p<0.1; ** p<0.05; *** p<0.01

Topic Sentiment Analysis - OLS

Table 27: Results from OLS of monetary sentiments score at t_1

	Dependent variable: ΔBOK_{t+1}						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ΔBOK_t	0.007	0.012	0.015	−0.013	0.005	0.012	−0.025**
	(0.655)	(1.195)	(1.417)	(−1.295)	(0.495)	(1.165)	(−2.549)
$\Delta(\pi_t - \pi^*)$		0.080**	0.080**	0.043	0.042	0.068**	0.064**
		(2.450)	(2.427)	(1.483)	(1.342)	(2.105)	(2.381)
$\Delta(y_t - y^*)$		2.194***	1.843***	1.670***	1.743***	1.895***	1.554***
		(4.530)	(3.656)	(3.839)	(3.782)	(3.833)	(3.748)
$\Delta\pi_t^e$			0.093	0.066	0.072	0.093	0.053
			(1.039)	(0.852)	(0.884)	(1.063)	(0.727)
Δy_t^e			0.097**	0.050	0.052	0.078*	0.091**
			(2.144)	(1.269)	(1.242)	(1.750)	(2.464)
$Tone_{lex}$				0.251***			
				(6.697)			
경제현황					0.508***		
					(5.099)		
대외여건						0.447**	
						(2.505)	
통화정책방향							0.662***
							(7.913)
Observations	132	132	132	132	132	132	132
Adjusted R^2	−0.004	0.164	0.184	0.394	0.319	0.217	0.452

Note: *p<0.1; **p<0.05; ***p<0.01

Topic Sentiment Analysis - OLS

Table 28: Results from OLS of monetary sentiments score at t_2

	Dependent variable: ΔBOK_{t+2}						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ΔBOK_t	−0.005	−0.001	0.0005	−0.022*	−0.008	−0.002	−0.030**
	(−0.466)	(−0.107)	(0.041)	(−1.941)	(−0.758)	(−0.150)	(−2.513)
$\Delta(\pi_t - \pi^*)$		0.064*	0.068*	0.037	0.035	0.059*	0.055*
		(1.845)	(1.910)	(1.103)	(1.004)	(1.667)	(1.701)
$\Delta(y_t - y^*)$		1.581***	1.525***	1.363***	1.424***	1.559***	1.291**
		(3.082)	(2.799)	(2.697)	(2.749)	(2.879)	(2.573)
$\Delta\pi_t^e$			−0.040	−0.060	−0.056	−0.039	−0.069
			(−0.414)	(−0.678)	(−0.613)	(−0.412)	(−0.778)
Δy_t^e			0.019	−0.018	−0.018	0.006	0.015
			(0.385)	(−0.394)	(−0.377)	(0.113)	(0.343)
$Tone_{lex}$				0.204***			
				(4.688)			
경제현황					0.427***		
					(3.830)		
대외여건						0.327*	
						(1.684)	
통화정책방향							0.501***
							(4.982)
Observations	131	131	131	131	131	131	131
Adjusted R^2	−0.006	0.077	0.066	0.200	0.158	0.079	0.215

Note: * p<0.1; ** p<0.05; *** p<0.01

Appendix: eKoNLPy Guide

eKoNLPy Guide: Installation

eKoNLPy를 사용하기 위해서는 먼저 KoNLPy[9]가 설치되어 있어야 함. 설치방법은 `http://konlpy.org/en/v0.4.4/install`을 참조. 추가로 반드시 MeCab을 설치하여야 한다.

```
> pip install konlpy
> bash <(curl -s https://raw.githubusercontent.com/konlpy/konlpy/master/scripts/
    mecab.sh)
```

Listing 1: installation for Mac OS

eKoNLPy는 github 소스를 복제하여 설치한다.

```
> git clone https://github.com/entelecheia/eKoNLPy.git
> cd eKoNLPy
> pip install .
```

Listing 2: eKoNLPy installation command

eKoNLPy Guide: Part of Speech Tagging

형태소 분석은 KoNLPy와 동일하게 Mecab.pos(phrase)를 사용합니다. 먼저 KoNLPy의 Mecab 형태소 분석기로 처리한 후, 템플릿에 등록된 연속된 토큰의 조합이 사용자 사전에 등록되어 있으면 복합명사로 어절을 분리한다.

```
from ekonlpy.tag import Mecab
mecab = Mecab()
tokens = mecab.pos("양호한 이머징 경기흐름을 바탕으로 유가 등 원자재 가격이 큰 폭으로 상승했다")
mecab.sent_words(tokens)
mecab.replace_synonyms(tokens)
mecab.lemmatize(tokens)
```

Listing 3: Part of speech tagging

동의어 처리와 lemmatisation은 각각 *replace_synonyms()*와 *lemmatize()*을 사용한다. *sent_words(phrase, replace_synonym=True, lemmatisation=True, exclude_terms=True, remove_tag=False)* 함수는 감성분석을 위해 동의어 처리, lemmatisation을 모두 거친 후 토큰 중에서 명사, 형용사, 부사, 동사 등 필요한 토큰만을 추려내는 기능을 제공한다.

eKoNLPy Guide: Dictionary Management

ekonlpy.tag의 Mecab은 *add_dictionary(words, tag, force=False)*를 통하여 str 혹은 list of str 형식의 단어를 사전에 추가할 수 있습니다. 태그 분류는 Table (29)를 참조. *add_synonym(word, synonym, tag='NNG')*, *add_lemma(word, lemma)* 함수를 통해 각각 동의어와 표제어를 추가할 수 있다.

```
mecab.add_dictionary("금통위", "NNG")
mecab.add_synonym("더블딥", "이중침체", tag="NNG")
mecab.add_lemma("커져야", "크")

mecab.load_dictionary("wordlist.txt", tag="NNG")
mecab.load_synonyms("synonyms.txt", tag="NNG")
```

Listing 4: Dictionary management

load_dictionary(fname, tag='NNG'), *load_synonym(fname, tag='NNG')* 함수를 사용하면 파일에서 한번에 사전에 추가할 수 있다. dictionary 파일은 한 줄에 한단어로 저장되어 있어야 한다. synonym 파일은 한 줄에 '단어 동의어' 순으로 정의되어 있어야 하며 두 단어는 공백으로 분리되어 있어야 한다.

eKoNLPy Guide: Sentiment Analysis

단어사전 방식의 통화정책 센티멘트 분석을 위해서는 ekonlpy.sentiment의 **MPKO** 클래스를 사용한다.

```
from ekonlpy.sentiment import MPKO
mpko = MPKO(kind=1)
tokens = mpko.tokenize(text)
score = mpko.get_score(tokens)
```

Listing 5: Moneytary policy sentiment analysis

MPKO는 현재 kind parameter를 통해 3가지의 사전을 선택할 수 있다.

```
0: a lexicon file generated using Naive-bayes classifier with 5-gram tokens as
    features and changes of call rates as positive/negative label.
1: a lexicon file generated by polarity induction and seed propagation method with
     5-gram tokens.
2: a lexicon file to measure uncertainty, which is generated by seed propagation
    method.
```

Classifier를 이용하여 통화정책 센티멘트를 분석하기 위해서는 ekonlpy.sentiment의 **MPCK** 클래스를 사용한다.

```
from ekonlpy.sentiment import MPCK
mpck = MPCK()
tokens = mpck.tokenize(text)
ngrams = mpck.ngramize(tokens)
score = mpck.classify(tokens + ngrams, intensity_cutoff=1.5)
```

Listing 6: Moneytary policy sentiment classifier

intensity_cutoff parameter를 사용해 분류정확도가 낮은 문장을 neutral로 분류하는 강도를 설정할 수 있다. (default: 1.3)

eKoNLPy Guide: Sentiment Analysis

ekonlpy.sentiment는 KSA, HIV4, LM 3개의 다른 sentiment 분석용 클래스를 제공한다. KSA는 일반적인 한국어 감성분석 용도, HIV4는 Harvard IV-4 dictionary를 사용하는 일반 영어 감성분석 용도, LM은 Loughran and McDonald dictionary를 사용하는 영어 금융 분야 감성분석 용도이다. 영어의 경우 NLTK tokenizer와 Porter Stemmer를 사용하여 tokenizing을 한다.

```
from ekonlpy.sentiment import KSA
ksa = KSA()
score = ksa.get_score(ksa.tokenize(text))
from ekonlpy.sentiment import HIV4
hiv = HIV4()
score = hiv.get_score(hiv.tokenize(text))
from ekonlpy.sentiment import LM
lm = LM()
score = lm.get_score(lm.tokenize(text))
```

Listing 7: Other sentiment analysis classes

eKoNLPy Guide: Topic Analysis

To analyze the Monetary Policy Topics, create an instance of the **MPTK** class in ekonlpy.topic

```
from ekonlpy.topic import MPTK
mptk = MPTK()
tokens = mptk.nouns(text)
bow = mpko.doc2bow(tokens)
dtm = mpko.get_document_topic(bow)
```

Listing 8: Topic analysis

parammeters for *get_document_topic* fucntion

```
include_names: If True, return tuples of list including topic names.
                    ex) (topic_id, topic_name, topic_weight)
               If False (default), return tuples of list without topic name.
                    ex) (topic_id, topic_weight)
min_weight: If min_weight is set, return topics with the topic weight is greather
    than the min_weight.
            Otherwise, return all available topics.
```

eKoNLPy Guide: Tagset

Table 29: Tagset used in Mecab tagger of eKoNLPy

Tag	Name	Tag	Name	Tag	Name	Tag	Name
NNG	일반 명사	VCN	부정 지정사	JKQ	인용격 조사	XSA	형용사 파생 접미사
NNP	고유 명사	MM	관형사	JC	접속 조사	XR	어근
NNB	의존 명사	MAG	일반 부사	JX	보조사	SF	마침표, 물음표, 느낌표
NNBC	단위를 나타내는 명사	MAJ	접속 부사	EP	선어말어미	SE	줄임표 …
NR	수사	IC	감탄사	EF	종결 어미	SSO	여는 괄호 (, [
NP	대명사	JKS	주격 조사	EC	연결 어미	SSC	닫는 괄호),]
VV	동사	JKC	보격 조사	ETN	명사형 전성 어미	SC	구분자 , · / :
VA	형용사	JKG	관형격 조사	ETM	관형형 전성 어미	SY	기타 기호
VAX	파생형용사	JKO	목적격 조사	XPN	체언 접두사	SH	한자
VX	보조 용언	JKB	부사격 조사	XSN	명사파생 접미사	SL	외국어
VCP	긍정 지정사	JKV	호격 조사	XSV	동사 파생 접미사	SN	숫자

eKoNLPy is Open Source Software, and is released under the license GPL v3.

References

[1] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022, 2003.

[2] Kenneth Ward Church and Patrick Hanks. Word association norms, mutual information, and lexicography. In *the 27th annual meeting*, pages 76–83, Morristown, NJ, USA, 1989. Association for Computational Linguistics.

[3] N Godbole, M Srinivasaiah, S Skiena Icwsm, and 2007. Large-Scale Sentiment Analysis for News and Blogs. *uvm.edu*.

[4] William L Hamilton, Kevin Clark, Jure Leskovec, and Dan Jurafsky. Inducing Domain-Specific Sentiment Lexicons from Unlabeled Corpora. *Proceedings of the Conference on Empirical Methods in Natural Language Processing. Conference on Empirical Methods in Natural Language Processing*, 2016:595–605, November 2016.

[5] Matthew Hoffman, Francis R Bach, and David M Blei. Online learning for latent dirichlet allocation. In *advances in neural information processing systems*, pages 856–864, 2010.

References ii

[6] N Jegadeesh and D Wu. Word power: A new approach for content analysis. *Journal of financial economics*, 2013.

[7] Dong-Joo Lee, Jong-Heum Yeon, In-Beom Hwang, and Sang-Goo Lee. Kkma: a tool for utilizing sejong corpus based on relational database. *Journal of KIISE: Computing Practices and Letters*, 16(11):1046–1050, 2010.

[8] T Loughran and B McDonald. When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *The Journal of Finance*, 2011.

[9] Eunjeong L Park and Sungzoon Cho. Konlpy: Korean natural language processing in python. In *Proceedings of the 26th Annual Conference on Human & Cognitive Language Technology*, pages 133–136, 2014.

[10] Zhe Zhao, Tao Liu, Shen Li, Bofang Li, and Xiaoyong Du. Ngram2vec: Learning Improved Word Representations from Ngram Co-occurrence Statistics. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 244–253, Stroudsburg, PA, USA, 2017. Association for Computational Linguistics.