Bank of Korea

NLP & Financial Analysis

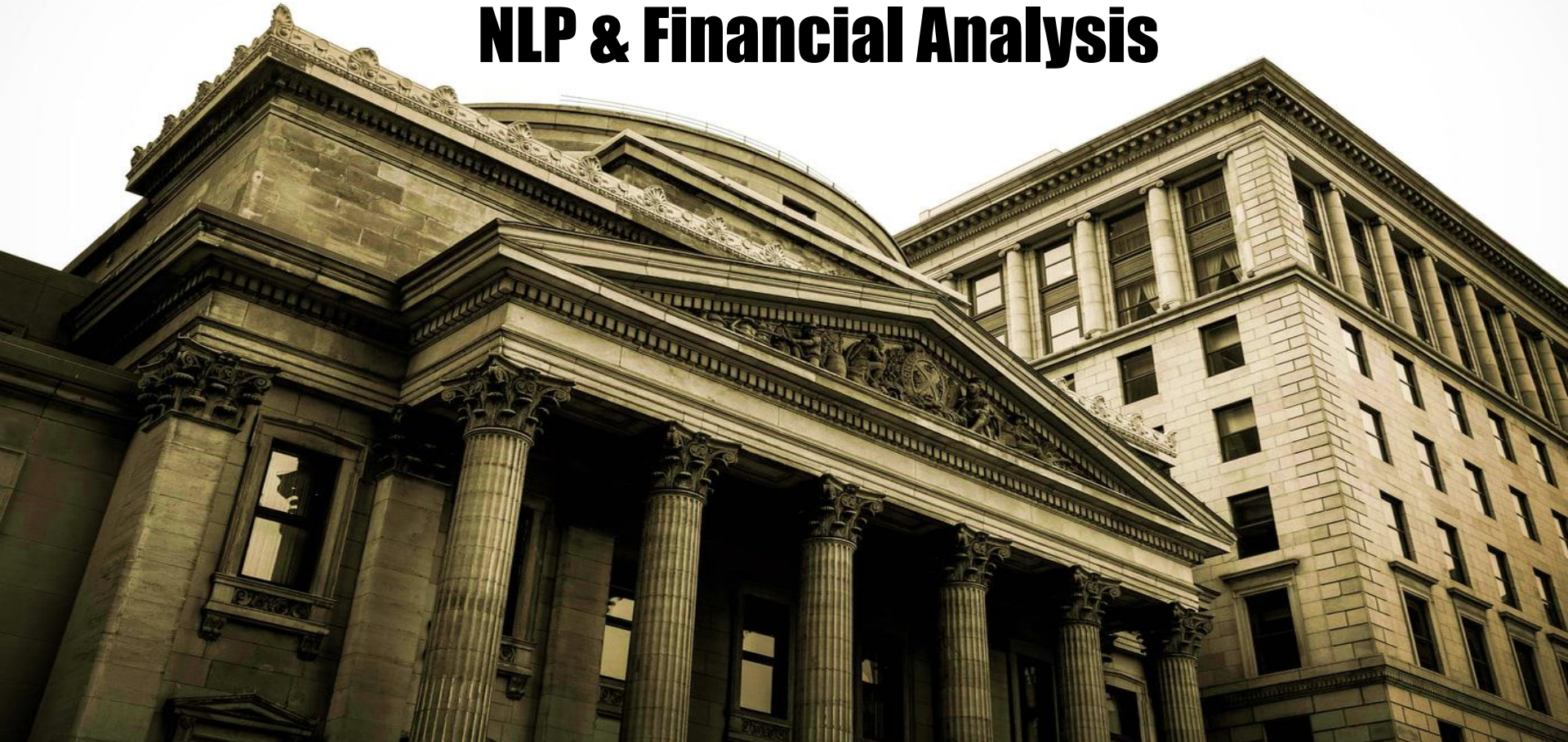


Table of Contents

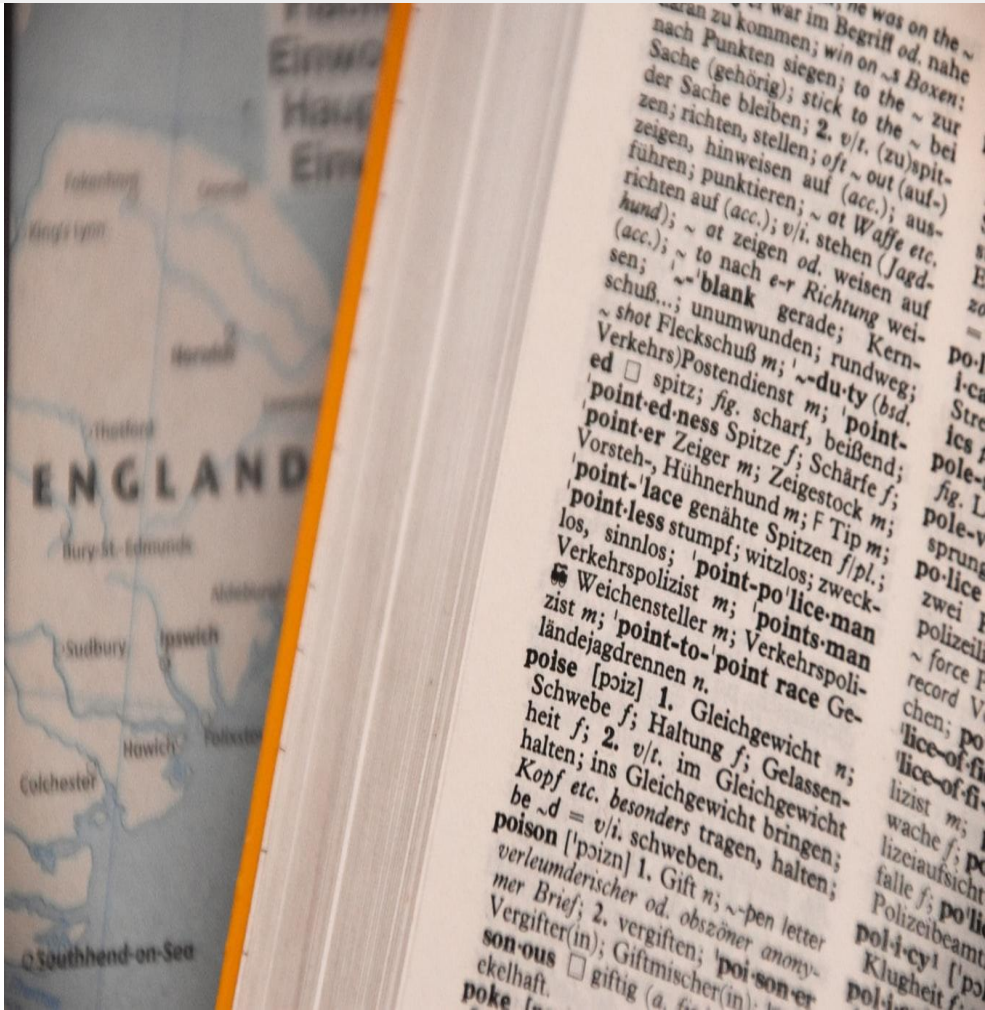
Abstract

Abstract

- **Text Mining**
  - MPB 회의록 분석
  - 기사 분석
  - 채권분석가 보고서

- **eKoNLPy 기법**
  - 경제전문용어
  - 동의어, 줄임말

- **Labeling**
  - 09’5 ~ 18’1’문서
  - 2,341 문장



Contribute

N-gram

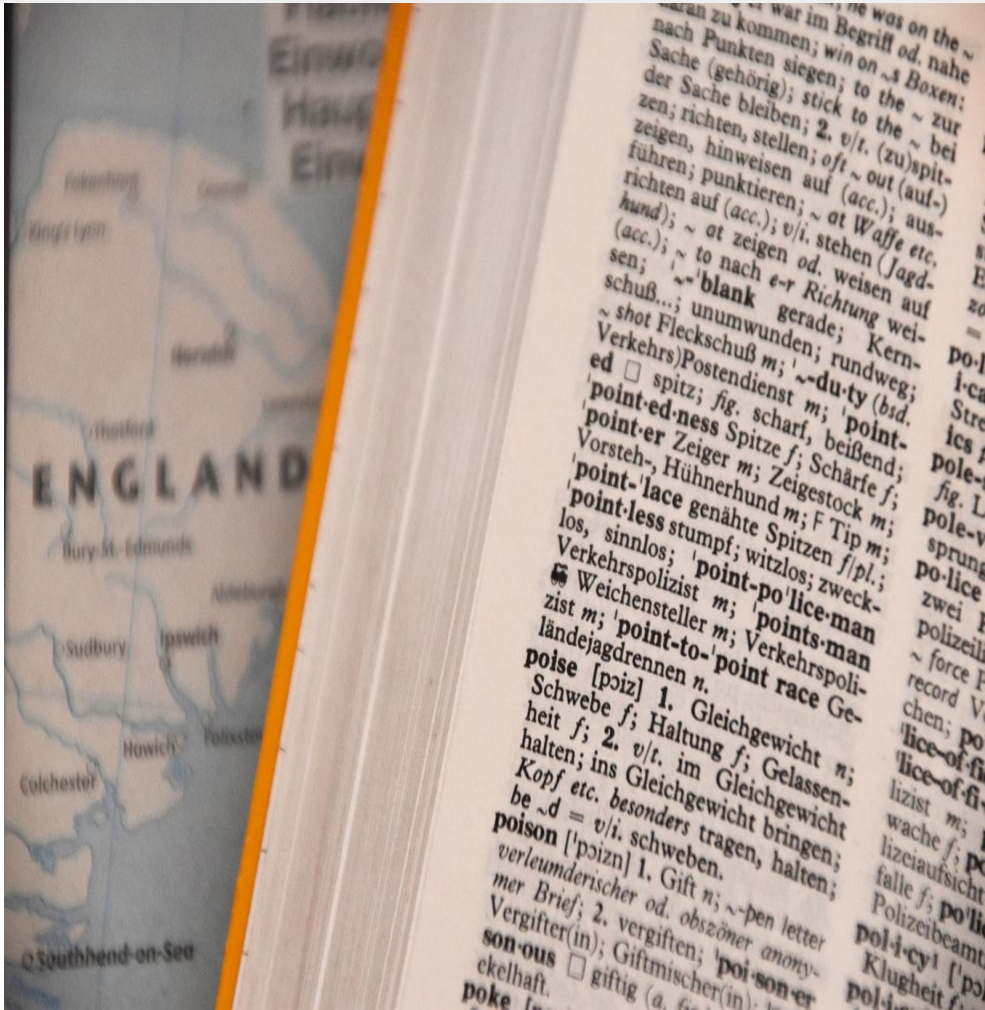
최초의 연

Contribute

eKoNLPy

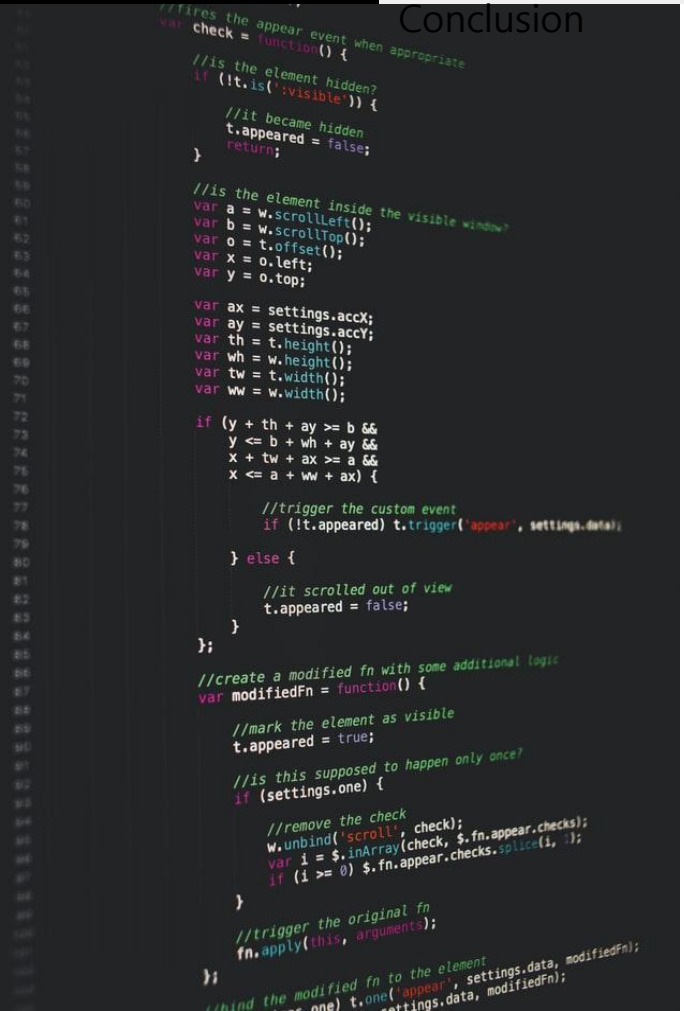
SentProp

Frameword



Data Sets

- MPB 회의록(151)
- 금리 기사(206,233)
- 채권분석가 보고서(26,284)

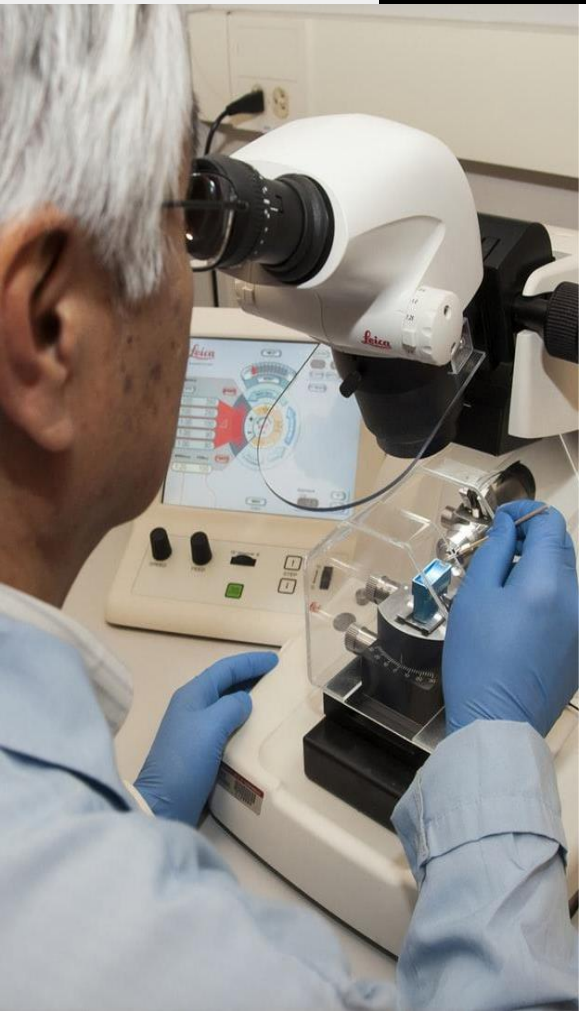


Data Processing

- **Tokenize & PoS Tagging(5개)**
  - 일반명사 부사 형용사 동사 부정어

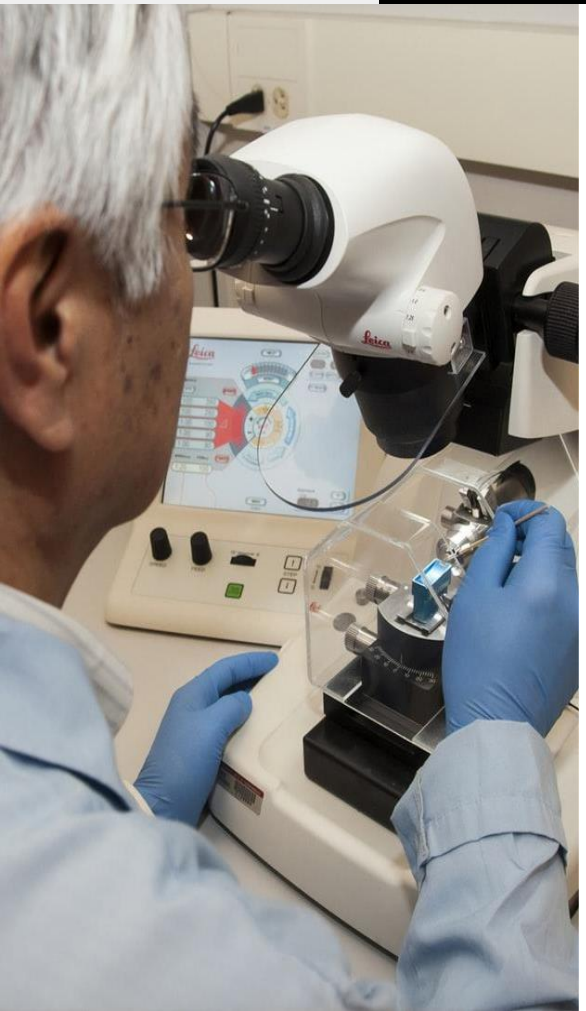
- **Normalizing**
  - **Stemming: 형식적인 어원추출**
    - having -> hav
  - **Lemmatization: 사전적인 어원추출**
    - having -> have

Analysis Methods



Analysis Methods

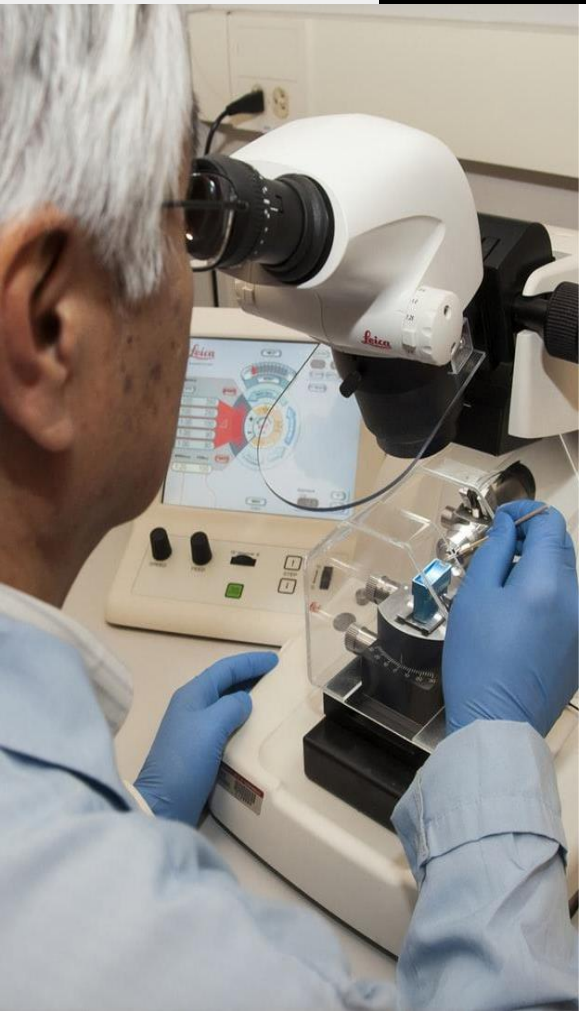
- **Why eKoNLPy?**
  - PostPosition / Space bar
  - Foreign language problem
  - Homonym
  - Irregular verb and adjective combination



Analysis Methods

- **Why eKoNLPy?**
  - PostPosition / Space bar
  - Foreign language problem
  - Homonym
  - Irregular verb and adjective combination

eKoNLPy

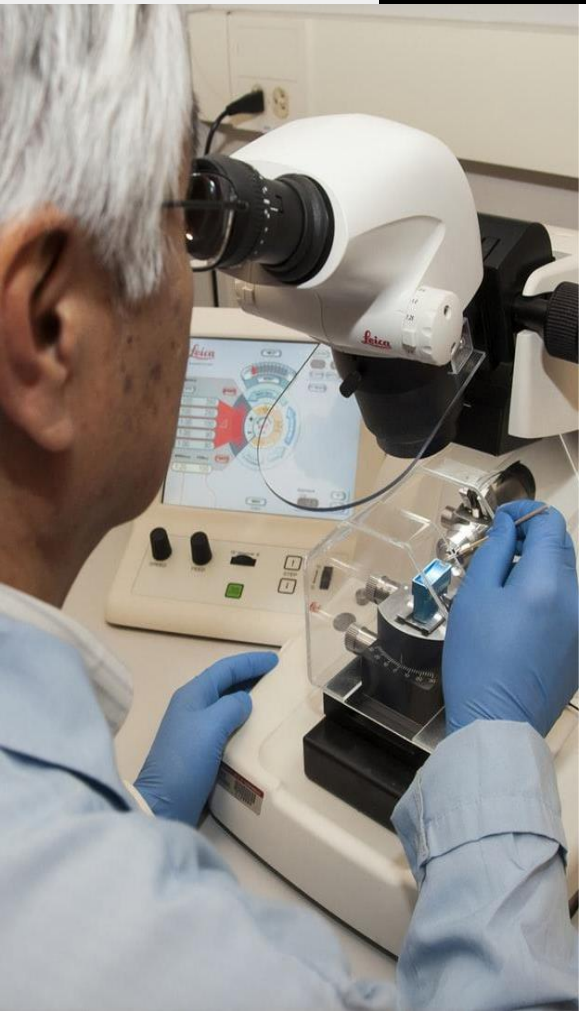


Analysis Methods

- **N-gram**
  - Rule 1 : 5-gram
  - Rule 2 : 15 under out

Processing ↑

MemorySize ↓



Analysis Methods

< 매파와 비둘기파의 정치적, 경제적 성향 비교 >

구 분	매파 (Hawkish)	비둘기파 (Dovish)
정치/외교적 성향	강경파	온건파
경제적 성향	▷ 물가안정 위주 (인플레이션 억제) ▷ 긴축정책과 금리인상을 주장 ▷ 경제적으로 진보성향	▷ 경제성장 위주 (인플레이션 장려) ▷ 양적완화와 금리인하를 주장 ▷ 경제적으로 보수성향

Analysis Methods
(how to distinguish dovish / hawkish)

1. Supervised vs Unsupervised
   - Google Cloud sentiment API — Supervised
   - PMI(Point-wise Mutual Information) — Unsupervised

1. Machine-learning-based vs Lexical-based

Market Approach

Manual

Dictionary based

corpus based

Analysis Methods
(Machine Learning based)

- **Market Approach**
  - **text** *Dependent variable*
  - **Economical** *Independent variable*
  - **Not Subjective judgement**

If the interest rate 1-month change is positive it is classified as hawkish and opposing.

Analysis Methods
(Lexical based)

- **Lexical Approach**
- **Intuitive**
- PMI(Point-wise Mutual Information)
- **Doesn’t recognize the antonym**
- **The result depends on seed-word**

Use Ngram2vec and Select SentProp

By bootstrapping the seed-word

발표자 변경
유정현 -> 이규호

Evaluation

- 평가 : Lexical
  - 검증용 데이터
  - 2341개(BOK introductory statements , 2009.5~2018.1)
- 위의 과정으로 통해 제작한 사전 > Naive Bayes 분류기로 분류

Evaluation

- dovish/hawkish 문장을 60(train),40(test)% 나눠 30번 반복
  - train_set을 통해 나온 결과 : 86%
- 나머지 40%의 테스트 데이터를 통해 검증해본 결과
  - Marketing = 68%
  - Lexical = 67%

Evaluation

- 서로 다른 관점의 두가지 접근법으로 만든 n-gram 사전이 좋은 유사도를 보인다(69%)
- 이 사전을 이용한 분석이 완전 새로운 문서분석에서도 좋은 결과를 보인다(68%, 67%)

Measuring Sentiments

1. 문장의 톤을 파악

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

1. 해당 문서의 톤을 파악

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

특징 : -1에 가까울 수록 dovish, 1에 가까울 수록 hawkish

marketing과 lexical 따로 한다

Measuring Sentiments

1. 문장의 톤을 파악

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

0 - 1
0 + 1

1. 해당 문서의 톤을 파악

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

1 - 0
1 + 0

특징 : -1에 가까울 수록 dovish, 1에 가까울 수록 hawkish

marketing과 lexical 따로 한다

Measuring Sentiments

- 새로운 문서를 Dovish / Hawkish로 NLP를 통해 분석할 수 있는 모델을 완성
- 현존하는 경제 지표들과 비교분석 가능

Empirical Analysis

1. 현재 및 미래의 통화 정책 결정을 기존 거시경제 데이터 보다 더욱 잘 설명할 수 있는가? -> (TOPIC1)

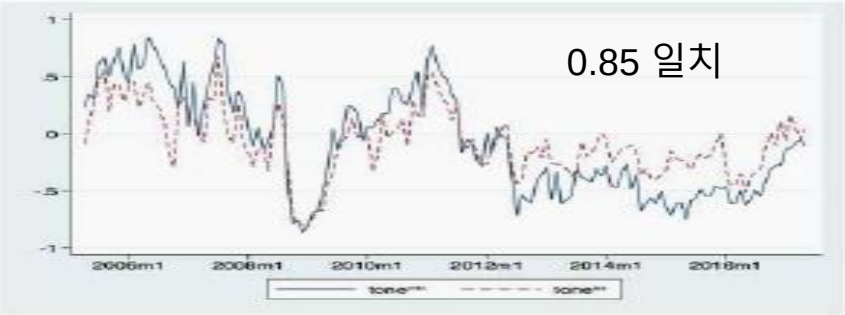
1. (경제)분야에 특화된 사전을 이용하는 것이 중요한가? -> (TOPIC2)

1. 한국어>영어로 번역 된 문서분석보다 한국어 원문 문서 분석이 더 나은가? -> (TOPIC3)

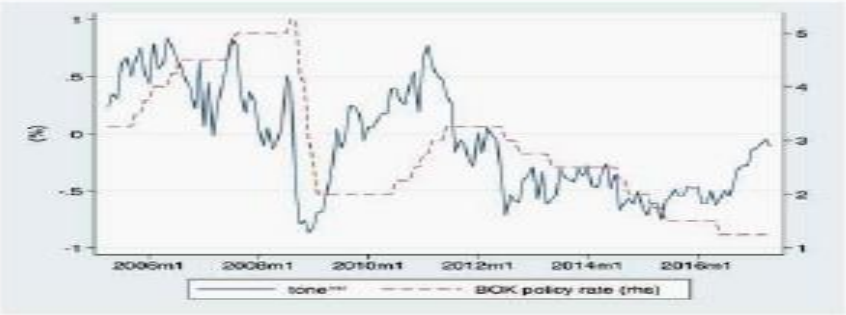
Empirical Analysis

(1. Measures of MP Sentiment)

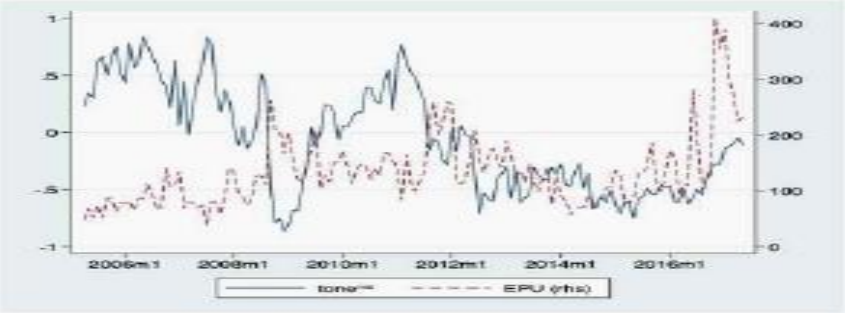
Figure 4. MP Sentiments, BOK Policy rate, and Other Measures of Uncertainty



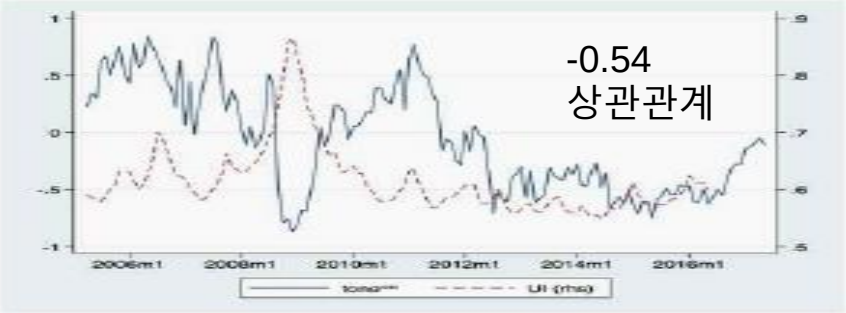
(a) $tone^{mkt}$ and $tone^{lex}$



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



(c) $tone^{mkt}$ and EPU

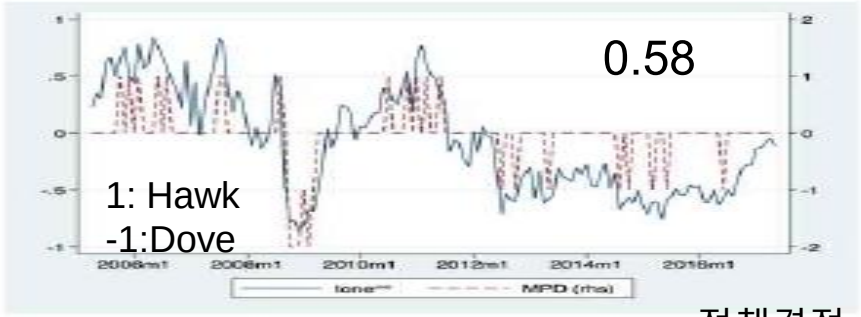


(d) $tone^{mkt}$ and UI

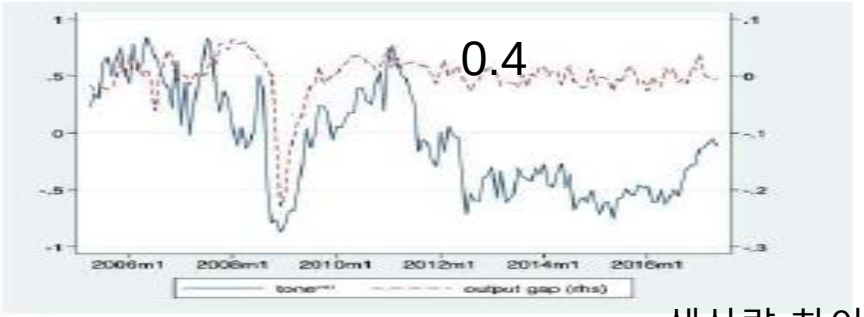
Empirical Analysis

(1. Measures of MP Sentiment)

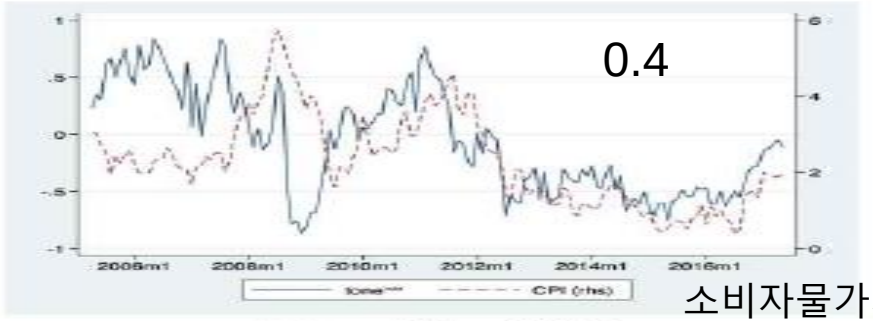
Figure 5. MP Sentiment and Macroeconomic Variables



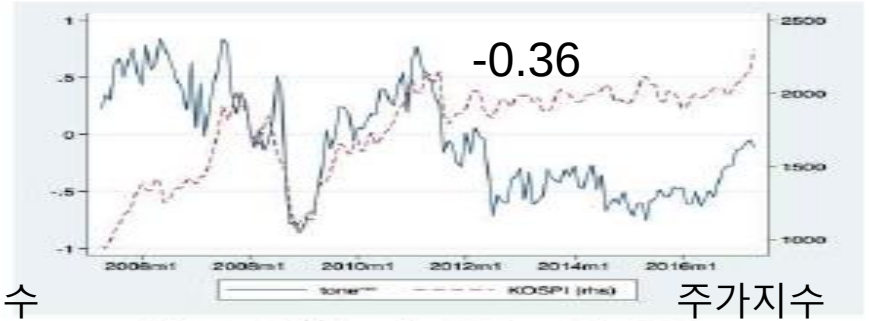
(a) $tone^{mkt}$ and MP decisions 정책결정



(b) $tone^{mkt}$ and output gap 생산량 차이



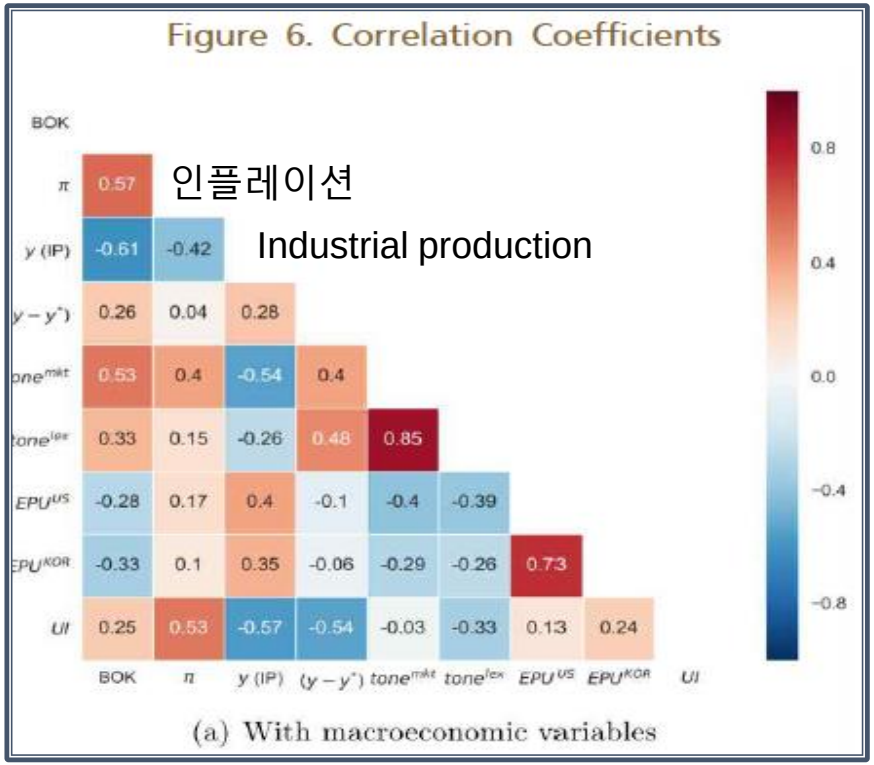
(c) $tone^{mkt}$ and CPI 소비자물가지수

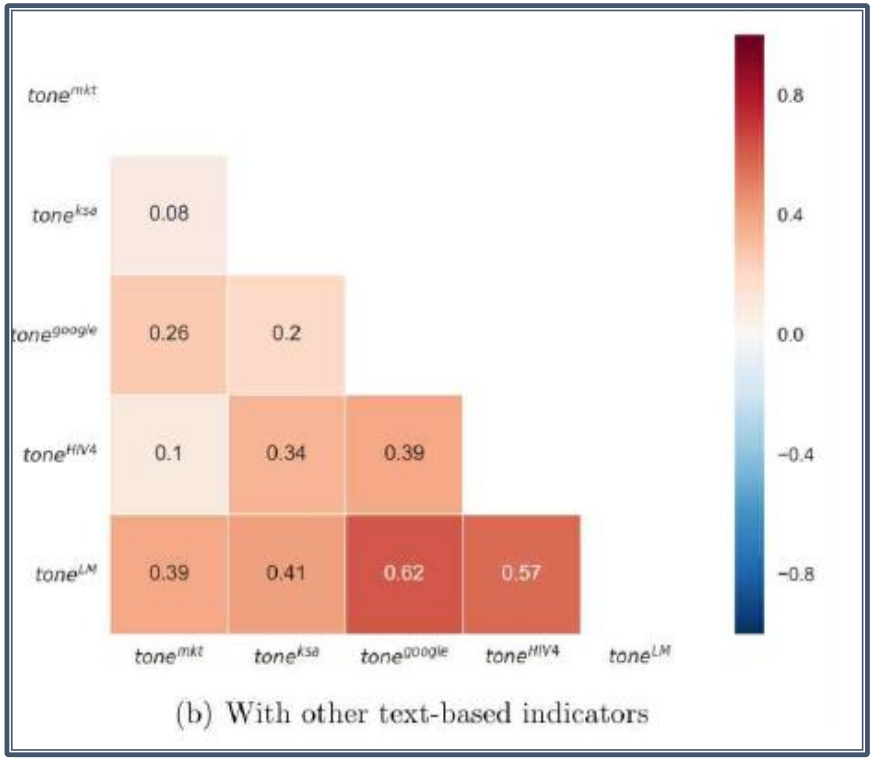


(d) $tone^{mkt}$ and stock market index 주가지수

Empirical Analysis

(1. Measures of MP Sentiment)





Empirical Analysis, (TOPIC1)

- 현재 및 미래의 통화 정책 결정을 기존 거시경제 데이터보다 더욱 잘 설명할 수 있는가?

MacroEconomic Model VS MacroEconomic Model + Lexicon

R^2	0.095	➔	0.446 (Table 7,현재)
R^2	0.109	➔	0.461 (Table 8,미래)
R^2	0.08	➔	0.37 (Apel & Grimaldi)

$$MP_t = \alpha + \rho MP_{t-1} + \gamma_1(\pi_t - \pi^*) + \gamma_2(y_t - y_t^*) + \gamma_3\pi_t^e + \gamma_4 y_t^e + \epsilon_t,$$

$$\Delta MP_t = \rho \Delta MP_{t-1} + \gamma_1 \Delta(\pi_t - \pi^*) + \gamma_2 \Delta(y_t - y_t^*) + \gamma_3 \Delta\pi_t^e + \gamma_4 \Delta y_t^e + \beta X_t + u_t,$$

미래예측은 t ➔ t+k, t-1 ➔ t

Empirical Analysis, (TOPIC1)

- 현재 및 미래의 통화 정책 결정을 기존 거시경제 데이터보다 더욱 잘 설명할 수 있는가?

MacroEconomic Model VS MacroEconomic Model + Lexicon

R^2 0.095 ➔ 0.446 (Table 7,현재)
R^2 0.109 ➔ 0.461 (Table 8,미래)
R^2 0.08 ➔ 0.37 (Apel & Grimaldi)

$$\underset{(0.65)}{\Delta \hat{r}_{t+1} = 1.90 \Delta r_t} + \underset{(4.64)}{7.28 IP growth_t} + \underset{(0.10)}{0.12 CPI_t}, \quad pseudo\, R^2 = 0.08.$$

With $tone_t^{mkt}$, we obtain

$$\underset{(0.90)}{\Delta \hat{r}_{t+1} = -1.67 \Delta r_t} + \underset{(0.86)}{4.20\, tone_t^{mkt}} + \underset{(5.33)}{9.73\, IP growth_t} - \underset{(0.14)}{0.28\, CPI_t},$$
$$pseudo\, R^2 = 0.37.$$

Empirical Analysis

- 해당 분야특화 사전을 이용하는 것이 중요한가?
- 한영 텍스트가 아닌 오리지날 한국어 텍스트를 사용하는 것이 유효한가?

- tone mkt -한국은행/한국어/경제분야특화 사전 분석기 0.446
- tone ksa -서울대/한국어/일반사전 분석기
- tone google -구글/영어/일반사전 분석기
- tone HIV4 -하버드/영어/일반사전 분석기
- tone LM -Loughran & McDonald/영어/경제분야특화 사전 분석기 0.127

Empirical Analysis

- 해당 분야특화 사전을 이용하는 것이 중요한가?
- 한영 텍스트가 아닌 오리지날 한국어 텍스트를 사용하는 것이 유효한가?

- tone mkt -한국은행/한국어/경제분야특화 사전 분석기 **0.446**

- tone ksa -서울대/한국어/일반사전 분석기

- tone google -구글/영어/일반사전 분석기

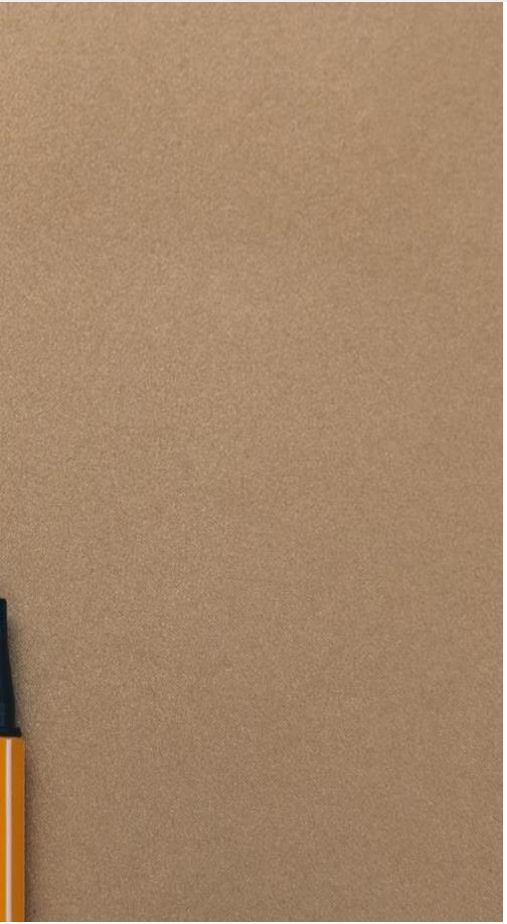
- tone HIV4 -하버드/영어/일반사전 분석기

- tone LM -Loughran & McDonald/영어/경제분야특화 사전 분석기 **0.127**

Empirical Analysis

- 해당 분야특화 사전을 이용하는 것이 중요한가?
- 한영 텍스트가 아닌 오리지날 한국어 텍스트를 사용하는 것이 유효한가?

- tone mkt -한국은행/한국어/경제분야특화 사전 분석기 **0.446**
- tone google -구글/영어/일반사전 분석기
- tone LM -Loughran & McDonald/영어/경제분야특화 사전 분석기 0.127



Conclusion

- 경제분야에서 (거의) 최초로 자연어처리와 감성분석을 통한 의미있는 분석방법을 제시함
- 영어번역/일반사전에 비해 한국어 원문/분야특화사전의 유용성을 밝힘
- 현재 거시경제지표를 통한 분석보다 유용함을 보여줌
- 중앙은행의 영향력에 대한 평가/예측/설명에 도움이 됨
- 다른분야에도 적용하기 쉬움

Suggestions & Questions

For English-based text analysis, we translate all the MPB’s minutes into English using **Google Cloud Translation**.42) measures the tone of minutes using the service of sentiment analysis provided by Google Cloud Natural Language.43) is based on the general-purpose **Harvard IV-4 dictionary** and is based on the **field-specific dictionary of Loughran and McDonald**(2011).

영어 기반 텍스트 분석의 경우 Google Cloud Translation을 사용하여 MPB의 모든 분을 영어로 번역합니다. Google Cloud Natural Language에서 제공하는 감정 분석 서비스를 사용하여 분의 톤을 측정하는 것은 범용 Harvard IV-4 사전을 기반으로하며 Loughran 및 McDonald (2011)의 필드 별 사전을 기반으로합니다.

	Positive precision	Positive recall	Negative precision	Negative recall	
Market approach	63	75	74	62	평균 68.5 **분산 36.25**
Lexical approach	69	71	65	62	평균 66.75 **분산 12.18**

QnA