KO R EA
컨퍼런스

납세자의 정서가 조세정책에 미치는 영향

KoNLP, N2H4

황의찬

2021-11-19

발표 개요

납세자의 조세저항은 정부의 조세정책결정에 영향을 주며 텍스트 데이터 분석에 관한 중요성이 증가하고 있다.

납세자의 조세저항을 측정하기 위해 감정지수를 산출하고 세법개정안을 중심으로 감정지수가 조세정책에 미치는 영향을 확인한다.

KoNLP, N2H4을 활용하여 댓글분석을 해보자.

1. 조세정책

2. 웹크롤링과 온톨로지

3. 감정지수

4. 마무리

2019년 신용카드 소득공제

- 신용카드 소득공제 제도는 제도의 일몰이 정해져 있는 조세특례제한법에 속함

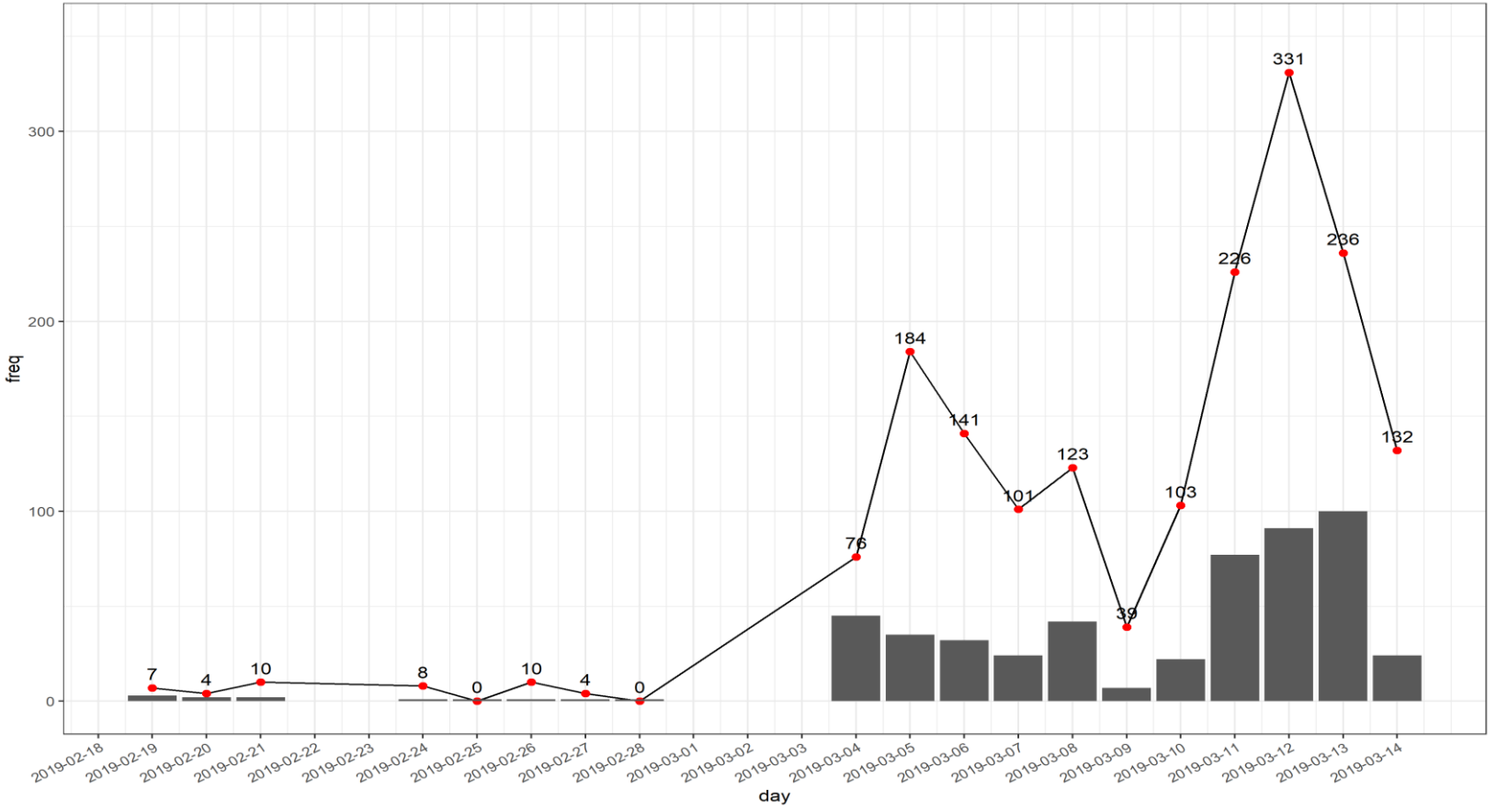
1. 홍남기 경제부총리 신용카드 소득공제율 축소 또는 일몰 암시
   2019-03-04 납세자의 날 행사 발언

2. 부총리 발언 관련 신용카드 소득공제 축소나 폐지 없음
   2019-03-11 기획재정부 입장 발표

3. 조특법 신용카드 소득공제율 유지 일몰 3년 연장 발표
   2019-03-13 당정청 협의회 발표

2019년 신용카드 소득공제

KO R EA 컨퍼런스



자료 선정과정

KO R EA 컨퍼런스

1. 네이버 뉴스검색 : 세법개정안
   - 1년 단위로 기간 설정
     예시) 2020.1.1 ~ 2020.12.31
     예시) 2020년의 경우 네이버뉴스 기사 2299건

2. 각 연도별 기사제목 명사추출하여 텍스트마이닝
   - 각 연도별 이슈가 되었던 세법개정안 확인

     예시) 2020년 세목관련 단어 빈도순위
     부동산(1), 종부세(9), 소득세(10), 양도세(11) 주식(13),
     대주주(20), 가상화폐(33), 전자담배(160)

자료

KOREA 컨퍼런스

세목	키워드	검색기간
2013 소득세	근로소득공제	2013.08.07 2014.01.01
2013 조특법	신용카드소득공제	2013.08.07 2013.12.31
2014 개별소비세	담배세	2014.09.10 2014.12.02
2014 지방세	지방세	2014.09.11 2014.12.08
2017 법인세	법인세	2017.08.01 2017.12.05
2017 소득세	소득세	2017.08.01 2017.12.05
2018 종합부동산세	종부세	2018.07.05 2018.12.08
2019 조특법	신용카드소득공제	2019.03.03 2019.12.10
2020 소득세	가상자산 소득세	2020.07.21 2020.12.02
2020 소득세/법인세/종부세	부동산 3법	2020.07.10 2020.08.04
2020 개별소비세	전자담배 개별소비세	2020.07.21 2020.11.30
2020 소득세	금융투자소득	2020.06.26 2020.12.31

자료

	뉴스	댓글	단어	감정지수
2013 소득세	473	704	8117	-55.88
2013 조특법	301	287	2989	-65.85
2014 개별소비세	138	338	3294	-50.91
2014 지방세	41	45	370	-82.56
2017 법인세	5010	273	2339	-44.91
2017 소득세	88	188	1788	-56.17
2018 종합부동산세	2370	10685	90141	-67.05
2019 조특법	1317	3760	26679	-75.04
2020 소득세	201	1087	7712	-82.62
2020 부동산3법	81	546	4479	-63.33
2020 개별소비세	239	974	7727	-96.63
2020 소득세	540	2814	21622	-69.87
	10799	21701	177257	

발표 개요

납세자의 조세저항은 정부의 조세정책결정에 영향을 주며 텍스트 데이터 분석에 관한 중요성이 증가하고 있다.

납세자의 조세저항을 측정하기 위해 감정지수를 산출하고 세법개정안을 중심으로 감정지수가 조세정책에 미치는 영향을 확인한다.

KoNLP, N2H4을 활용하여 댓글분석을 해보자.

1. 조세정책

2. 웹크롤링과 온톨로지

3. 감정지수

4. 마무리

조세정책 온톨로지

KO R EA 컨퍼런스

1) 정책타임라인 기준 검색기간 설정이후 네이버 뉴스 검색 2) 뉴스댓글 추출	1) 댓글 형태소 분해하여 문장 뜻 형성에 핵심적 역할하는 용언,체언만 추출 2) 출현빈도 40번 이상 용언,체언 필터링	1) 용언,체언에 매칭된 감성단어 필터링 이후 감성점수 태깅 2) 조세 온톨로지에 범용 온톨로지 결합
1단계:웹크롤링	2단계:형태소분석	3단계:온톨로지구축

웹크롤링 코드

KO R EA 컨퍼런스

```
#1 근로소득공제----
# start = 정책발표일, end = 정책철회or통과일
start_chr <- "2013.08.07"
end_chr <- "2014.01.01"

# 큰 그릇 생성(네이버 기사 링크 담을 그릇)
url_list <- c()
url_list <- as.data.frame(url_list)

for (i in seq(1,5001,10)) {

  #페이지별 루프 실행
  paste0("https://search.naver.com/search.naver?&where=news&query
=%EA%B7%BC%EB%A1%9C%EC%86%8C%EB%93%9D%EA%B3%B5%EC%A0%9C&sm=tab_pge&sort=0&photo=0&field
=0&reporter_article=&pd=3&ds=",start_chr,"&de=",end_chr,"&docid=&nso=so:r,p:from20130807to20140101
,a:all&mynews=0&cluster_rank=16&start=",i,"&refresh_start=0") -> page

  #페이지 내 url수집
  read_html(page) %>%
    html_nodes(".info_group a") %>%
    html_attr("href") -> url

  #url 중 네이버뉴스형식 제공되는 것만 필터링
  urls <- url[grep("news.naver.com",url)]
  print(length(urls))
  urls <- as.data.frame(urls)
  url_list <- bind_rows(urls, url_list)

}
# 중복 url 삭제
url_list %>%
  unique() -> url_list
```

웹크롤링 코드

KO R EA 컨퍼런스

```r
# 기사와 댓글 담을 그릇 준비
contents <- c()
comments <- c()

# url별 기사 및 댓글 수집 loop

for (i in 1:nrow(url_list)) {
  content <- c()
  url <- url_list[i,]
  content <- getContent(url)
  comment <- getComment(url)

  contents <- bind_rows(content, contents)
  comments <- bind_rows(comment, comments)

  print(i)
}

# comments$contents --> spacing by Kospacing
comments$contents <- spacing(comments$contents)
comments$contents <- unlist(comments$contents)
```

조세정책 온톨로지 예시

KO R EA 컨퍼런스

	word	score
117	가르치며	1
118	가르침	1
119	가르침을 받아	0
120	가리지 못하다	-1
121	가리지 않고	0
122	가마득하다	-1
123	가만있지 못하고	-1
124	가망 없는	-1
125	가망이 없게	-2
126	가벼운 마음으로	1
127	가벼운 말이나	-1
128	가벼운 상처	-1
129	가벼운 슬픔	-1
130	가벼운 행동으로	-1
131	가볍게	-1
132	가볍게 까부는	-1
133	가볍게 까불다	-1
134	가볍게 쑤시듯	-1
135	가볍게 웃다	1
136	가볍게 채신없이	-1
137	가볍게 행동하거나	-1
138	가볍게 행동하는	-1
139	가볍게 행동하다	-1
140	가볍고	1
141	가볍고 보드랍게	1
142	가볍고 상쾌하다	2
143	가볍고 상쾌한	2
144	가볍고 시원하게	2
145	가볍고 편안하게	2
146	가볍고 환하게	2
147	가분가분	1
148	가분히	1
149	가뿐가뿐	1

Showing 116 to 150 of 16,577 entries, 2 total columns

	word	score
5761	부끄러운 느낌이	-1
5762	부끄러운 부분	-1
5763	부끄러운 줄	-1
5764	부끄러운 줄 모르고	-1
5765	부끄러운 짓을	-1
5766	부끄러움	-1
5767	부끄러움을	-1
5768	부끄러움을 느끼게	-1
5769	부끄러워서	-1
5770	부끄러워지다	-1
5771	부끄러워하는	-1
5772	부끄러워하는 마음	-1
5773	부끄러워하다	-1
5774	부끄러워함	-1
5775	부끄러이	-1
5776	부끄럽게	-1
5777	부끄럽게 여기다	-1
5778	부끄럽다	-1
5779	부담	-2
5780	부담스러운	-2
5781	부담스럽다	-2
5782	부담이 없이	2
5783	부당하거나	-2
5784	부당하게	-2
5785	부당하게 가로채어	-2
5786	부당하다	-1
5787	부당한	-1
5788	부도덕	-2
5789	부도덕한	-1
5790	부드러우며	1
5791	부드러우면서	1
5792	부드러우면서 아삭하다	1
5793	부드러운	1

Showing 5,760 to 5,794 of 16,577 entries, 2 total columns

	word	score
16354	퍼주니가없지	-1
16355	퍼줄라고	-1
16356	퍼줄라구	-1
16357	폐지하려면	-1
16358	폐지해라	-2
16359	폔해지지	-1
16360	폐지하는	-1
16361	폐지하는건지	1
16362	폐지하라는	-1
16363	폐지하려니	-1
16364	폐지하면	-1
16365	폐지하자	-1
16366	폐지하지	1
16367	폐지한다니요	-1
16368	폐지해야	0
16369	폐지해야한다	1
16370	폐진하는건	-1
16371	포로정권	1
16372	폭망하거나	-1
16373	폭탄세금땜에	-1
16374	표받을려고하는	-1
16375	풀어야할	-1
16376	풋내기	-1
16377	필요없고	0
16378	필요없는	-1
16379	필요없다	-1
16380	필요없다고	-1
16381	필요없다그냥	-1
16382	필요없다는거네요	-1
16383	필요없으면버리고	-1
16384	필요없지요	-1
16385	필요한정책이면	1
16386	필요해	-1

Showing 16,353 to 16,387 of 16,577 entries, 2 total columns

	word	score
16537	해줍니다	-1
16538	해줘도	1
16539	했다며	-1
16540	했다면	1
16541	행정부	-1
16542	허접하게	-1
16543	헛돈이나	-1
16544	현금쓰고	-1
16545	현금쓰면	-1
16546	현금쓰세요	-1
16547	현금쓰지	-1
16548	현금없으면	-1
16549	현금없이	-1
16550	현금이없었어	-1
16551	현실화되고있구나	-1
16552	현정권	-1
16553	현정권의	-1
16554	현정부	-1
16555	현정부가	-1
16556	현정부는	-1
16557	현정부만	1
16558	현정부의	-1
16559	현찰이있어야	-1
16560	혈세돈만	-1
16561	혜택없는	-1
16562	혜택이있지	0
16563	효구로만드는	-1
16564	혼밥하시고	-1
16565	확대하고	1
16566	확인하기	-1
16567	환급하는거	-1
16568	환효하고에휴나라꼴이	-1
16569	후련하구만	-1

Showing 16,536 to 16,570 of 16,577 entries, 2 total columns

발표 개요

납세자의 조세저항은 정부의 조세정책결정에 영향을 주며 텍스트 데이터 분석에 관한 중요성이 증가하고 있다.

납세자의 조세저항을 측정하기 위해 감정지수를 산출하고 세법개정안을 중심으로 감정지수가 조세정책에 미치는 영향을 확인한다.

KoNLP, N2H4을 활용하여 댓글분석을 해보자.

1. 조세정책

2. 웹크롤링과 온톨로지

3. 감정지수

4. 마무리

감정지수 산출과정

KO R EA 컨퍼런스

* 감성지수를 계산하여 조세정책결정에 미치는 영향 분석

1. 감성지수
   - 네이버 댓글 웹크롤링
   - 텍스트마이닝을 통하여 구축한 조세정책 온톨로지 활용하여 감성지수 산출

$$Sentiment\ \ Index = \frac{\sum_{j=1}^{n} sentiment_{score}}{\sum_{j=1}^{n} count(word)}$$

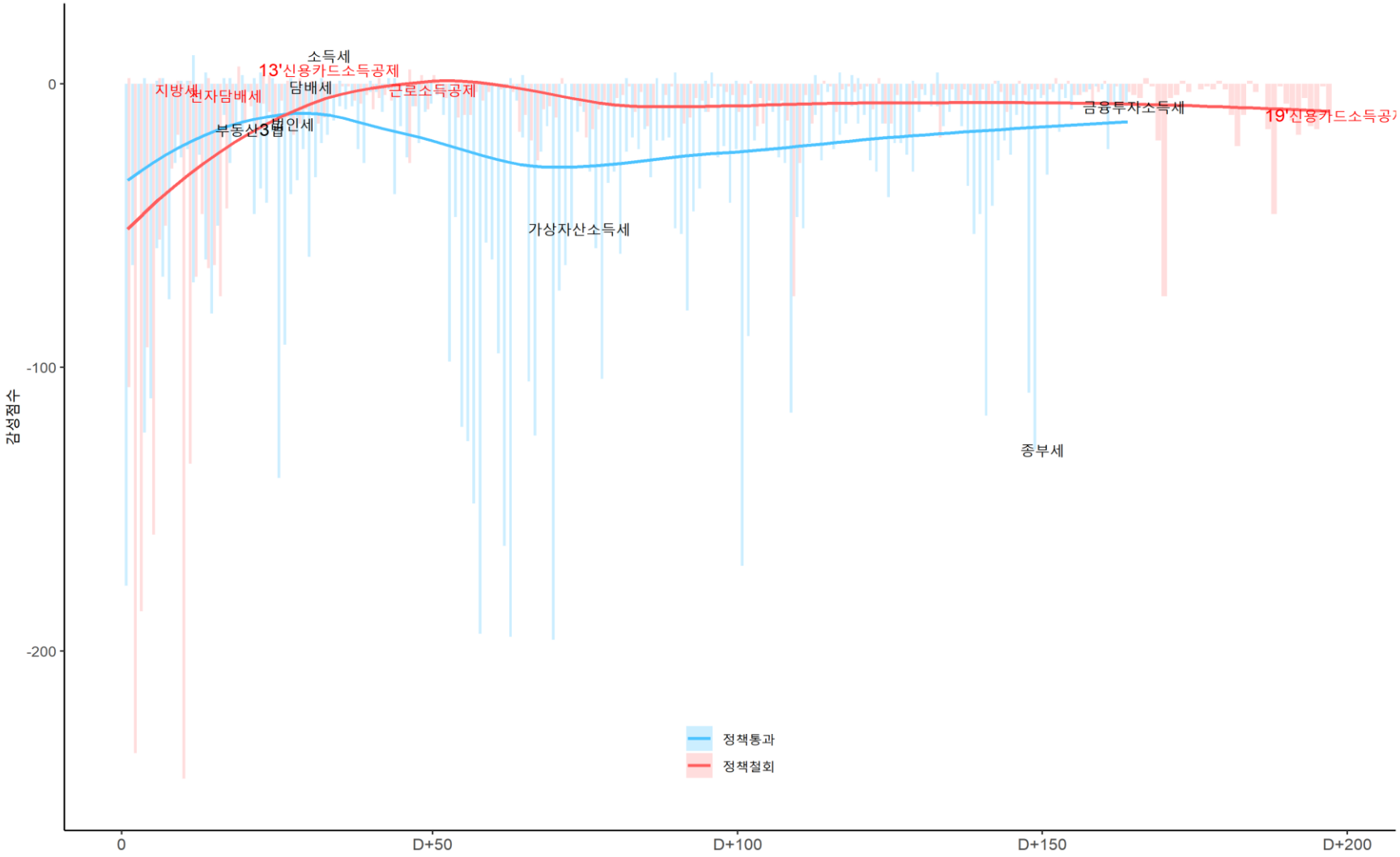
2. 정책결정
   - 정책통과: 세법개정안 가결
   - 정책철회: 세법개정안 폐기 / 핵심적 내용 수정된 경우

감정지수 코드


```
# 감정지수 산출
textmining3 %>%
  mutate(index = sentiment/freq) %>%
  dplyr::group_by(regTime) -> textmining3

textmining3 %>%
  dplyr::group_by(regTime) %>%
  summarise(aggregate_index = sum(sentiment) / sum(freq)) -> temp

textmining4 <- bind_rows(textmining4, temp)
# 단어 수 계산
textmining %>%
  filter(regTime %in% datebreaks[[i]]) %>%
  group_by(regTime) %>%
  dplyr::summarize(freq = n()) -> textmining_count
# 온톨로지와 결합
textmining %>%
  inner_join(ontology, by = "word") -> textmining
textmining <- textmining[,c(3,2,1,4)]

  #감성점수 기준 그룹화
  textmining %>%
    filter(regTime %in% datebreaks[[i]]) %>%
    group_by(regTime) %>%
    dplyr::summarise(sentiment = sum(score)) -> textmining2
```

정책결정과 감정지수

KO R EA
컨퍼런스



주1: X축: 정책타임라인, Y축: 감성점수

주2: 정책철회한 케이스의 경우 정책통과한 케이스에 비해 초기 감성점수가 낮은 경향이 있다.

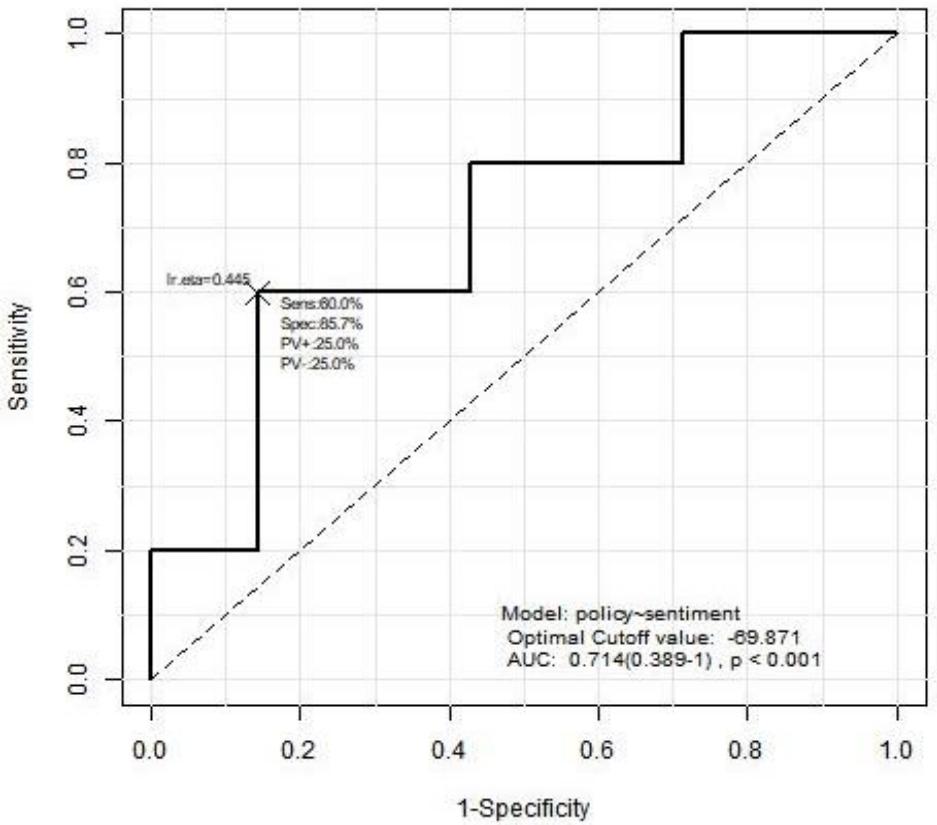
분석결과

	Dependent variable policy		
	OLS	logistic	AME
	(1)	(2)	(3)
index	-0.016	-0.076	-0.015
	(0.01)	(0.054)	(0.007)
			[-0.027,-0.002]
Constant	-0.641	-5.54	
	(0.674)	(3.739)	
Z			-2.209
P-value			0.027
Observations	12	12	
R2	0.205		
Adjusted R2	0.125		
Log Likelihood		-6.827	
Akaike Inf. Crit.		17.655	
Residual Std. Error	0.482		

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

AME: Average Marginal Effect

ROC-curve



X축: 1-특이도(FPR), Y축: 민감도(TPR)

임계치별로 민감도 특이도를 계산하여 roc-curve로 나타내고 민감도+특이도가 최대가 되는 최적임계치 도출

1. optimal_cutoff 0.445
2. optimal_cutpoint -69.871
3. AUC(Area under the curve) 0.714

AUC는 ROC-curve 아래 면적으로 분류기의 분류 성능을 의미함

주1: X축: 정책타임라인, Y축: 감성점수
주2: 정책철회한 케이스의 경우 정책통과한 케이스에 비해 초기 감성점수가 낮은 경향이 있다.

결론

1. 조세정책 온톨로지를 구축하고 세법개정안에 대한 감성지수를 측정한다.

2. 감성지수와 조세정책결정의 연관성을 확인한다. 분석결과 조세저항이 컸던 조세정책에서 감성지수가 1표준편차 낮아질때 정책철회할 확률은 증가한다.

3. 조세정책에 대한 감성지수 -69.87을 기준으로 분류했을때 분류정확도는 75%로 가장 높았다.

발표 개요

납세자의 조세저항은 정부의 조세정책결정에 영향을 주며 텍스트 데이터 분석에 관한 중요성이 증가하고 있다.

납세자의 조세저항을 측정하기 위해 감정지수를 산출하고 세법개정안을 중심으로 감정지수가 조세정책에 미치는 영향을 확인한다.

KoNLP, N2H4을 활용하여 댓글분석을 해보자.

1. 조세정책

2. 웹크롤링과 온톨로지

3. 감정지수

4. 마무리

경청해 주셔서
감사합니다.

R 사용자회