

조세재정 Brief

K I P F I S S U E P A P E R

텍스트 분석을 이용한 조세정책에 대한 인식 연구

배진수 한국조세재정연구원 부연구위원 / jsbae@kipf.re.kr

- 01 연구 배경
- 02 분석 내용
- 03 결론 및 시사점
참고문헌

2024. 5. 14.

No.167



kipf 한국조세재정연구원
KOREA INSTITUTE OF PUBLIC FINANCE

요약

- 본고에서는 텍스트 분석 방법론을 이용해 인터넷 미디어 여론이 조세정책에 관한 국민들의 선호를 잘 반영하는지에 대해 정성적·정량적인 분석을 시도하였음
 - 구체적으로, 설문을 통해 가상자산소득 기타소득세 신설에 관한 의견을 텍스트 형태로 수집하였으며 이를 인터넷 미디어 여론의 텍스트와 비교 분석함
- 텍스트를 구성한 주제의 비율을 분석하기 위해 잠재 디리클레 할당(Latent Dirichlet Allocation, LDA)을 이용한 토픽모형 분석을 하였으며 다음과 같은 결과를 얻음
 - 가상자산소득 기타소득세에 대한 설문응답을 토픽모형을 이용해 분석하였을 때에는 수익에 대한 과세, 불로소득에 대한 과세, 투자자 보호, 과도한 세율 등 조세정책에 대한 의견들이 주요 주제를 구성하였으나,
 - 해당 세제 신설에 대한 인터넷 미디어 여론을 분석한 경우 해당 정책에 대한 주제도 발견할 수 있었지만, 부동산 정책, 정부에 대한 불만, 금융투자소득세, 선거 등 주제가 다양하고 산발적인 경향을 보였음
- 해당 세제 신설에 대한 긍정/부정 의견 비율을 비교하기 위해 트랜스포머 모델을 이용한 감성분석을 수행하였으며 다음과 같은 결과를 얻음
 - 인터넷 미디어 여론에서 가상자산소득 기타소득세 신설에 대해 부정 감성으로 분류되는 비율이 설문에서 얻은 부정 응답의 비율에 비해 현저하게 높은 것으로 나타남
 - 이러한 차이가 나타나는 원인은 해당 정책에 대해 부정적인 의견을 가진 사람들이 편향적으로 인터넷 미디어 여론을 조성하는 것으로 분석됨

※ 본 조세재정브리프는 “배진수·박정흠·김수현, 『텍스트 분석을 이용한 조세정책에 대한 인식 연구』, 한국조세재정연구원, 2022.” 중 일부 내용을 발췌·요약한 것임

01 연구 배경

- 국내의 조세정책 수립 과정은 국민들의 의견이 다양한 통로를 통해 반영될 수 있는 환경으로 변화하고 있음
 - 과거 조세정책의 수립은 정부가 발의하는 세법개정안을 중심으로 이루어졌으나 법안실명제가 도입된 2000년 이후 의원발의안의 수가 증가하고 있으며 2010년 이후에는 이러한 현상이 가속화되고 있음 (국회예산정책처, 2017)
- 한편 인터넷 미디어가 발달함에 따라 국민 개인도 조세정책에 관해 자유롭게 자신의 의견을 표출할 수 있는 여론형성 여건이 갖추어짐
- 조세정책 수립에 영향을 주는 다양한 형태의 여론들이 국민들의 실제 선호를 잘 반영하고 있는지를 평가하는 것은 중요한 의미를 지님

- 인터넷 미디어 여론 등을 통한 여러 형태의 의견수렴은 정부의 독단적인 정책결정을 견제하고 정책수립 과정에서 사회적 합의를 달성하는 데 기여할 수도 있으나,
 - 민주적 절차를 통해 권한을 위임받은 대의기관의 결의가 편향되거나 왜곡된 여론의 영향을 받게 된다면 이는 공정한 사회적 합의의 과정을 방해할 수도 있음
- 본 연구는 최근 발달하고 있는 텍스트 분석 방법론을 이용하여 인터넷 미디어 여론이 조세정책에 관한 국민들의 선호를 얼마나 잘 반영하는지에 대한 분석을 시도해 보았다는 점에서 의의가 있음
- 인터넷 미디어 여론은 뉴스 댓글이나, 청와대 국민 청원, 소셜 미디어 의견 게재 등 텍스트의 형태로 나타나기 때문에 수치화하기 어려운 비정형 데이터의 성격을 가지고 있음
 - 최근 발달하고 있는 텍스트 분석 방법론은 비정형 데이터를 수량화하여 분석할 수 있는 유용한 도구가 되고 있으며 미디어의 편향성 등을 연구하는 데 사용되고 있음
 - 본 연구에서는 인터넷 미디어 여론과 무작위 표본 추출을 통한 설문응답을 비교하기 위해 텍스트 분석 방법론을 사용함

02 분석 내용

1. 잠재 디리클레 할당(LDA)을 활용한 토픽모형 분석

- 우선 본 연구에서는 가상자산소득 기타소득세 신설에 관한 설문응답과 해당 정책에 관한 인터넷 뉴스 댓글을 수집¹⁾하여 그 주제의 구성비를 토픽모형을 통해 비교 분석하였음
 - 설문은 무작위 표본 추출을 통해 이루어졌으며 해당 정책에 대한 긍정/부정 의견을 텍스트로 수집하였음
 - 인터넷 뉴스 댓글은 가상자산 과세를 다루는 뉴스의 댓글을 수집하되 종합부동산세와 가상자산 규제 등의 키워드와 관련된 뉴스는 제외하여 수집 대상을 한정하였음
- 토픽모형 분석은 텍스트의 주제를 요약하고 분류하는 방법론이며, 잠재 디리클레 할당(LDA)은 어떤 주제에 특정 키워드가 나타날 조건부 확률을 통해 주제를 추정하는 토픽모형임
- LDA 토픽모형을 통하여 분석한 결과 설문응답을 구성하고 있는 주제들과 인터넷 뉴스 댓글을 구성하고 있는 주제들이 상이하게 나타나는 것을 확인할 수 있음(〈표 1〉 참조)
 - 설문을 통해 얻은 응답들에는 해당 조세정책의 도입과 밀접한 연관이 있는 주제들이 다수를 차지하는 것을 확인할 수 있었음
 - 세제 도입을 찬성하는 의견 중 가장 높은 비율(20%)로 나타난 주제 묶음은 '화폐, 돈, 수익, 사

1) 인터넷 뉴스 댓글 수집은 국내 최대 규모의 뉴스 유통 플랫폼인 네이버를 통해서 수행함

- 람, 가치, 이득, 자금'이라는 키워드를 포함하여 수익에 대한 과세의 당위성이라는 주제로 추정됨
- 두 번째로 높은 비율(15%)을 나타낸 주제는 '무분별, 투자, 규제, 범죄, 방지, 조세, 제도, 형평'을 포함하고 있어 무분별한 투자 방지를 위한 조세정책의 필요성이라는 주제로 판단됨
- 세 번째 주제는 불로소득 과세의 당위성 및 소득세와의 형평성, 네 번째 주제는 주식이나 다른 재산 과세와의 형평성을 의미하는 것으로 보임
- 한편 세제 도입을 반대하는 주제 중 가장 높은 비

- 율(20%)을 차지하는 주제는 '위험, 투자, 돈, 수단, 과세, 부당, 생각, 가치, 때문, 정도'로 구성되어 있으며 이는 위험을 감수한 투자소득에 대한 과세의 부당성을 언급하는 것으로 보임
- 두 번째로 높은 비율을 보인 주제(19%)는 '세금, 부담, 과도, 국가, 수익, 부분'을 포함하여 과도한 세율이라는 주제로 보이며 세 번째로 높은 비율을 나타낸 주제는 '가상, 화폐, 법, 인정, 손해, 안정'을 포함하여 법 제도로 인정되지 않은 가상자산 과세의 부당함과 관련된 주제로 나타남

표 1 조세정책 관련 개방형 설문에 따른 긍정/부정 응답 토픽 분석

긍정/부정	키워드	관련 응답 수(개)	비율
긍정 응답	화폐, 돈, 수익, 사람, 가치, 이득, 자금	271	0.20
	무분별, 투자, 규제, 투기, 범죄, 방지, 조세, 제도, 형평	209	0.15
	불로소득, 일반, 개인, 소득세, 세율, 적용, 원칙, 동일	195	0.14
	주식, 재산, 경우, 세금, 부과, 해당, 이유, 부여, 금액	186	0.14
	생각, 긍정, 정부, 국민, 공평, 현금, 부분, 일	171	0.13
	과세, 소득, 필요, 경제, 가능, 일종, 정책	168	0.12
	거래, 이익, 발생, 탈세, 때문, 국가, 과세, 사용, 인정	164	0.12
		(1,364)	(1.00)
부정 응답	위험, 투자, 돈, 수단, 과세, 부당, 생각, 가치, 때문, 정도	126	0.20
	세금, 부과, 과도, 국가, 수익, 부분	120	0.19
	가상, 화폐, 법, 인정, 손해, 안정	84	0.14
	가상, 자산, 실물, 시장, 실제 현금, 변동, 소득세	83	0.13
	세율, 때문, 부정, 적용, 개인, 부담, 이익, 발생	82	0.13
	투자자, 손실, 보호, 제도, 과세, 정부, 규제, 정책	73	0.12
	소득, 자산, 가상, 주식, 거래, 금융, 경우, 기준, 형평	49	0.08
		(617)	(1.00)

주: 키워드는 최대 10개로 제한하였으며, () 안의 숫자는 합계임
 자료: 배진수 외(2022), p. 105, <표 V-4>

표 2 가상자산소득 과세 관련 댓글 토픽모델 분석 결과

연도	키워드	댓글 수(개)	댓글 비율
2018	나라, 투자, 말, 국민, 부동산, 돈	867	0.44
	화폐, 돈, 가상, 코인, 사람, 과세, 주식, 거래소, 규제, 우리	572	0.29
	세금, 정부, 생각, 인정, 서민, 거래소, 폐쇄	552	0.28
		(1,991)	(1.00)
2019	주식, 나라, 니겟, 정권, 생각, 투자, 경제, 시장	801	0.35
	세금, 돈, 정부, 사람, 국민, 부과, 개	784	0.34
	세금, 화폐, 인정, 비트코인, 과세, 가상, 소득, 코인	697	0.31
		(2,282)	(1.00)
2020	주식, 나라, 시장, 개미, 과세, 양도세, 금융, 거래세, 우리	776	0.22
	세금, 말, 민주당, 하나, 한국, 개돼지	752	0.21
	화폐, 가상, 거래소, 코인, 인정, 자산, 암호	751	0.21
	정부, 국민, 정권, 사람, 세금, 개, 소득	624	0.18
	생각, 부동산, 정책, 투자, 비트코인, 경제, 국가	602	0.17
		(3,505)	(1.00)
2021	나라, 국민, 민주당, 정권, 선거, 때, 표, 정책, 내년, 대선	8,660	0.29
	정부, 생각, 말	5,876	0.20
	세금, 화폐, 가상, 보호, 정부, 자산, 인정, 국가	5,355	0.18
	사람, 비트코인, 코인, 부동산, 투기, 시장, 도박	4,938	0.17
	코인, 주식, 과세, 투자, 거래소, 소득	4,562	0.16
		(29,391)	(1.00)
2022	가상, 비트코인, 화폐, 과세	926	0.43
	세금, 주식, 공약, 폐지, 투자, 양도세, 공매도, 개	645	0.30
	코인, 것, 사람, 돈, 나라, 국민, 이재명, 대통령, 민주당, 윤석열	589	0.27
		(2,160)	(1.00)

주: 키워드는 10개 이하로 제한하였으며, () 안의 숫자는 합계임
 자료: 배진수 외(2022), p. 103, <표 V-3>

- 한편 인터넷 뉴스 댓글을 토픽모형으로 분석한 결과는 해당 세제와 관련된 주제뿐만 아니라 다른 주제들을 다수 포함하고 있었으며 주제가 다양하고 산발적인 특징을 보여주었음(〈표 2〉 참조)
 - 설문문의 부정응답 관련 주제 중 하나를 구성하고 있던 가상자산의 제도권 인정에 관한 주제는 빈번하게 등장하고 있음(2019년 주제3, 2020년 주제3, 2021년 주제3)
 - 또한 긍정응답 관련 주제 중 하나였던 무분별한 투기 방지를 위한 조세정책의 필요성 관련 주제도 발견할 수 있음(2021년 주제4)
 - 하지만 가상자산소득 기타소득세와 거리가 있는 주제들도 산발적으로 발견되고 있음
 - 댓글의 수집 과정에서 종합부동산세 관련 뉴스는 제외했음에도 불구하고 2018년에는 부동산 관련 주제(주제1)가 높은 비율(44%)을 차지하는 것을 확인할 수 있음
 - 또한 '개미, 공매도'와 같이 금융투자소득세 관련 키워드(2020년 주제1)나 선거와 관련된 주제(2022년 주제2)도 발견할 수 있음

2. 트랜스포머 모형을 이용한 감성 분석

- 이어서 본 연구에서는 가상자산소득 기타소득세 신설에 관한 인터넷 뉴스 댓글을 감성분석하여 그 결과를 설문응답에서 얻은 긍정/부정 비율과 비교하였음
 - 만일 인터넷 미디어 여론이 조세정책에 대한 실제 국민들의 선호를 잘 반영한다면 감성분석의 결과는 설문응답에서 얻은 긍정/부정 비율과 유사할 것으로 기대할 수 있음
 - 감성분석을 수행하기 위해서는 텍스트를 분석하여 긍정/부정의 감성을 분류해줄 감성분류기의 학습이 필요함
 - 본 연구에서는 구글이 개발하여 배포한 트랜스포

- 머 모형인 BERT(Bidirectional Encoder Representations from Transformers)를 응용하였음
 - 구체적으로 한글 자료로 사전학습이 되어 있는 KR-BERT를 본 연구에서 수행한 설문문의 내용으로 추가 학습하여 감성분류기를 구축함
- 트랜스포머 모형을 통해 구축된 감성분류기는 설문응답의 긍정/부정 감성을 높은 확률로 정확히 분류하였으며 인터넷 뉴스 댓글의 부정 감성도 잘 분류하였으나, 긍정 감성 댓글의 경우 절반 정도만 긍정으로 분류하였음
 - 세제 신설에 긍정적인 의견을 가진 댓글을 부정으로 분류한 주요 원인은 비속어 등 부정적 감정이 느껴지는 단어를 같이 사용하였거나, 부정에 대한 부정으로 긍정 의견을 표현하였거나, 의견이 명시적으로 드러나지 않고 문맥을 통해서만 해석할 수 있는 경우들이었음
 - 이후의 감성분석 결과는 감성분류기가 긍정 댓글을 부정으로 분류하였을 확률을 고려하여 긍정으로 분류된 댓글이 실제로는 배가 더 있을 수 있다는 점을 감안하여 해석함

- 인터넷 뉴스 댓글을 감성분석한 결과는 〈표 3〉과 같음
- 2018년부터 2022년까지 가상자산소득 기타소득세 신설에 긍정적인 감성으로 분류된 인터넷 뉴스 댓글은 약 8% 해당함
 - 이는 본 연구에서 사용된 감성분류기가 긍정 댓글을 약 절반 정도만 긍정으로 분류한다는 한계를 반영하더라도 전체 댓글의 16% 정도만이 세제 신설에 긍정적인 감성으로 분류될 수 있다는 것을 의미함
 - 반면 설문조사에서 세제 신설에 긍정적인 응답의 비율은 69.3%에 달하여 인터넷 뉴스 댓글의 긍정 분류 비율과 큰 차이를 보였음
 - 이러한 차이는 인터넷 미디어 여론이 조세정책에 대한 국민의 선호를 잘 반영하기보다는 편향되게 표출하고 있을 가능성을 시사함

표 3 연도별 인터넷 뉴스 댓글의 감성분석 결과

(단위: 개)

연도	단순합			가중치 고려		
	긍정	부정	합	긍정	부정	합
2018	171 (0.09)	1,826 (0.91)	1,997 (1.00)	1,587 (0.04)	37,060 (0.96)	38,647 (1.00)
2019	186 (0.08)	2,101 (0.92)	2,287 (1.00)	669 (0.04)	18,417 (0.96)	19,086 (1.00)
2020	275 (0.08)	3,241 (0.92)	3,516 (1.00)	680 (0.03)	24,804 (0.97)	25,484 (1.00)
2021	2,404 (0.08)	27,225 (0.92)	29,629 (1.00)	7,489 (0.04)	174,932 (0.96)	182,421 (1.00)
2022	166 (0.07)	2,184 (0.93)	2,350 (1.00)	680 (0.05)	12,674 (0.95)	13,354 (1.00)
계	3,202 (0.08)	36,577 (0.92)	39,779 (1.00)	11,105 (0.04)	267,887 (0.96)	278,992 (1.00)

주: 1. 가중치는 댓글을 읽은 다른 독자들이 댓글에 표시한 ‘좋아요’, ‘싫어요’의 개수
 2. () 안의 숫자는 비율

자료: 배진수 외(2022), p. 123, <표 V-10 >

- 인터넷 뉴스 댓글에 대해 독자들이 표시하는 ‘좋아요’와 ‘싫어요’ 수치를 가중치로 사용²⁾하여 감성분석을 시도하는 경우 부정적으로 분류되는 댓글의 비중이 더욱 높아짐
 - 이는 부정적인 의견을 가진 이용자들이 인터넷 미디어 여론에 더 많이 참여하고 있거나 혹은 더 자주 참여하고 있기 때문일 수 있음
 - 이러한 현상은 인터넷 미디어 여론이 국민들의 실제 선호를 편향되게 표출할 수 있는 기제로 작용할 수 있음

- 한편 인터넷 뉴스 댓글이 아닌 유튜브 뉴스 영상에 기록된 댓글을 감성분석하는 경우도 유사한 결과를 보여주고 있음
 - 가상자산 기타소득세 신설과 관련된 유튜브 뉴스 영상에 달린 댓글의 경우 13.3%가 긍정 감성으로 분류되었음
 - 이는 감성분류기가 참 긍정을 긍정으로 분류할 확률이 절반 정도라는 점을 감안해도 긍정 댓글 비율이 최대 27%를 넘지 않는다는 것을 의미함
 - 이는 본 연구의 결과가 플랫폼의 특성에 따른 특수한 결과가 아님을 보여주고 있음

2) ‘좋아요’와 ‘싫어요’는 댓글에 동의하거나 동의하지 않는 독자의 수를 나타내므로 ‘좋아요’를 댓글과 동일한 의견의 개수로 간주하고 ‘싫어요’를 댓글의 반대 의견의 개수로 간주함

03 결론 및 시사점

- 가상자산 기타소득세 신설에 관한 설문조사 응답과 인터넷 미디어 여론을 텍스트 분석 방법론으로 비교해 본 결과 인터넷 미디어 여론은 설문응답에 비해 다양하고 산발적인 주제를 포함하며 부정적인 감성을 표출하는 비율이 높음
 - LDA 토픽모형을 통해 분석한 결과 인터넷 미디어 여론은 해당 정책과 직접적으로 관련된 주제뿐만 아니라 일반적인 정부정책이나 정치 관련 주제들을 다양하게 포함하고 있음
 - 이는 정부정책이나 정치에 대한 일반적 견해가 특정 조세정책으로 전이되어 나타날 수 있는 환경을 조성함
 - 트랜스포머 모형을 통한 감성분석 결과는 인터넷 미디어 여론이 해당 세제 신설에 관해 설문응답보다 현저히 부정적인 감성을 나타내고 있음을 보여줌
 - 가중치를 통한 분석에 기초하면 이는 해당 정책에 부정적인 의견을 가진 사람들이 인터넷 여론에 더 많이 참여하거나 혹은 더 자주 참여하는 사람인 것으로 추정됨
 - 미디어 여론의 편향적 특성은 플랫폼의 특성과 무관하게 공통적으로 발견되는 현상임
- 이러한 결과는 인터넷 미디어 여론이 조세정책에 대한 국민들의 선호를 잘 반영하지 못할 가능성을 시사하고 있으며 조세정책 수립 과정에서 인터넷 미디어 여론을 수렴하는 경우 해석에 주의가 필요함
 - 인터넷 미디어 여론을 의식하여 대의기관이 수립한 조세정책을 수정하거나 철회하는 것은 불공정한 사회적 합의의 결과로 이어질 수 있으며 세제 안정성을 저해할 수 있음
 - 인터넷 미디어 여론을 정책 수립 과정에 반영하는 경우에도 그 특성을 감안하여 가능한 국민들의 실제 선호를 잘 반영할 수 있는 방법을 모색해야 함
 - 정부에 대한 단기적인 지지 성향이 장기적인 조세정책의 수립 과정에 전이되어 영향을 주는 것에 주의해야 하고,
 - 여론을 조성하는 이들의 구성이 편향되지 않은지 확인해야 하며 소수의 의견이 중복되어 다수의 의견처럼 보이고 있지 않은지도 주의해야 함

참고문헌

- 국회예산정책처, 『트렌드 세법: 과거 20년간 세법개정안의 궤적을 담다』, 2017.
- 배진수·박정흠·김수현, 『텍스트 분석을 이용한 조세정책에 대한 인식 연구』, 한국조세재정연구원, 2022.