



텍스트 분석을 이용한 조세정책에 대한 인식 연구

2022. 12.

배진수 · 박정흠 · 김수현



텍스트 분석을 이용한 조세정책에 대한 인식 연구

2022. 12.

배진수 · 박정흠 · 김수현

서 언

조세는 정부의 재정수요를 충족하기 위하여 국민에게 반대급부 없이 강제로 부과·징수하는 과징금으로서 국민의 재산권에 대한 침해적 성격이 있으므로, 그 과세요건은 반드시 국민의 대표기관인 국회가 제정하는 법률로 명확히 규정해야 한다. 이러한 원칙에 따라 조세정책은 필연적으로 세법개정안의 제출과 같은 정치적 입법과정을 통해 입안되며, 이 과정에서 국민의 선호가 잘 반영될 수 있도록 사회적 합의를 이루는 것이 중요하다고 할 수 있다.

과거 조세정책은 정부가 발의하는 세법개정안을 중심으로 이루어졌으나, 법안실명제가 도입된 2000년 이후 의원발의안의 수가 증가하기 시작했으며 2010년도 이후에는 이러한 현상이 가속화되고 있다. 한편 인터넷 미디어의 발달로 개인들도 조세정책에 관해 자유롭게 자신의 의견을 표출할 수 있는 여론 형성 여건이 갖추어졌다. 이러한 변화는 정부의 독단적인 정책 결정을 견제하고 조세정책이라는 정치적 과정에서 민의가 반영되고 사회적 합의를 달성할 수 있는 환경이 갖추어진 것으로 평가할 수 있다. 하지만 한편으로는 정치 양극화와 여론 양극화가 심화하고 있는 최근의 정치적 상황에서는 오히려 이념적으로 양극단에 있는 사람들의 의견들이 조세정책에 과도하게 반영되어 세계 안정성을 해칠 수 있는 환경을 조성한다고 할 수도 있다.

본고에서는 텍스트 분석 방법론을 이용하여 인터넷 미디어 여론이 조세정책에 관한 국민들의 선호를 얼마나 잘 반영하는지 정성적·정량적으로 분석하였고, 이를 통해 인터넷 미디어 여론의 조성이 조세정책에 대한 사회적 합의를 이루는 데 어떠한 의미를 지니는지 평가해 보았다. 구체적으로 본고에서는 설문문을 통해 가상자산소득 기타소득세 신설에 대한 의견을 텍스트 형태로 수집하였으며, 동시에 해당 정책에 대한 인터넷 뉴스 기사, 댓글을 수집하여 설문과 인터넷 미디어의 텍스트를 비교 분석하였다. 텍스트 분석

방법론을 사용한 분석 결과, 인터넷 미디어 여론의 경우는 가상자산소득 기타소득세 신설에 관한 직접적 의견을 표출하는 텍스트도 일부 있었으나, 집권 정당에 대한 일반적인 불만, 정부의 다른 정책에 대한 비판, 특정 정치인들에 대한 의견 등과 같이 조세정책과 관련이 없는 텍스트들도 많이 발견되어 조세정책의 의견을 수렴하기 위한 목적으로는 잡음이 심한 텍스트인 것으로 판단하였다. 또한 가상자산소득 기타소득세 신설에 관한 인터넷 미디어 여론에 대해 긍·부정의 감성분석을 수행한 결과, 부정의 감성이 설문 응답에서 얻어진 부정 의견보다 훨씬 높게 나타나 인터넷 미디어 여론은 편향성이 심한 것으로 나타났다.

조세정책에 대한 인터넷 미디어 여론의 편향성을 발견한 본 연구의 결과는, 조세정책의 수립과정에서 다양한 의견대립 속에서 국민들의 선호를 올바르게 도출하고 통합할 수 있는 과정과 능력이 더욱 중요해지고 있다는 시사점을 제공한다. 특히 기존의 선행연구들에서 발견된 인터넷 미디어 여론이 조세정책 의사결정에 영향을 준다는 결과와 정치 성향의 분포상 양극단에 있는 사람들이 온라인의 여론 형성에 적극 참여하고 있다는 결과와 합쳐볼 때, 본 연구의 결과는 인터넷 미디어 여론을 조세정책 수립과정에 여과 없이 반영하는 것에 대해서 경계해야 할 것이라는 메시지를 제공한다. 다만 본 연구는 특정 조세정책에 중심을 맞춰서 수행한 기초연구로서 본 연구 결과를 일반화하기 위해서는 향후 좀 더 분석의 범위를 넓히고 다양한 방법론을 이용한 발전된 기초연구, 정책연구, 학술적 연구가 지속적으로 수행되어야 할 것이다.

본 보고서는 한국조세재정연구원의 배진수 부연구위원, 박정흠 부연구위원, 전남대학교 김수현 교수가 공동으로 수행하였다. 저자들은 본 연구가 실험적인 기초연구임에도 불구하고 예산과 인력 등 여러 방면으로 아낌없이 지원해 준 한국조세재정연구원에 깊은 감사의 뜻을 표한다. 그리고 본 연구의 방향 수립과 완성도 향상에 많은 도움을 준 강동익, 김정환 부연구위원 등 연구질 관리 TF의 여러 구성원들과 외부 익명 심사의견을 준 두 분의 전문가에게 큰 감사의 마음을 전하고자 한다. 한편 자료 수집과 원고 정리의 과정에서 많은 도움을 준 원내의 윤다솜 연구원, 이아름 연구원, 최예나

연구원에게 감사의 뜻을 표한다.

마지막으로 본 연구의 결과는 저자들의 개인적인 학술적 견해이며 한국조세재정연구원의 공식적인 견해를 대변하는 것이 아님을 밝힌다.

2022년 12월

한국조세재정연구원
원장 김 재 진

요약 및 정책적 시사점

납세자의 여론이 조세정책의 수립과정에 영향을 줄 것이라는 사회적 인지도에도 불구하고 여론의 형성이 정책과정에 미치는 영향을 수치화한 연구는 발견하기 어렵다(황의찬·우석진, 2022). 이는 납세자들의 여론 형성은 인터넷 뉴스 댓글이라든가, 청와대 국민청원, 소셜 미디어(social media) 등과 같이 텍스트의 형태로 나타나기 때문에 수치화하기 어려운 성격이 있기 때문이라고 할 수 있다.

최근 발달하고 있는 텍스트 마이닝 기법은 뉴스나 댓글과 같은 텍스트 정보를 수량화된 배열로 나타내어 분석하는 이론적 틀을 제시하여 텍스트를 정량적으로 분석할 수 있는 기반을 제공하고 있다. 이러한 텍스트 분석 방법론을 이용하여 황의찬·우석진(2022)은 조세정책에 대한 인터넷 미디어 여론의 부정적인 감성이 강해질수록 정책이 수정되거나 철회될 가능성이 높다는 것을 보여주었다. 만일 인터넷 미디어 여론이 조세정책에 대한 국민들의 선호를 잘 반영한다면 이러한 현상은 바람직할 수 있지만, 만일 그렇지 못한 경우 인터넷 여론을 의식하여 정책을 수정하거나 철회하는 것은 사회적 합의를 반복하는 것으로 공정하지 못하며, 세제 안정성을 해칠 수도 있다.

본 연구에서는 Gentzkow et al.(2019)의 이론적 틀을 바탕으로 인터넷 미디어 여론이 조세정책에 대한 국민적 선호를 잘 반영하고 있는지를 텍스트 마이닝 방법론을 통해 정성적·정량적으로 분석해 보았다. 연구의 범위는 인터넷 미디어에서 활발하게 논의된 적이 있는 가상자산 기타소득세의 신설에 관한 여론을 분석하는 것을 중심으로 이루어졌다. 먼저 개방형 주관식 설문조사를 통해서 가상자산 기타소득세 신설에 관한 긍·부정 의견을 텍스트 형태로 수집하였다. 이와 동시에 해당 정책에 대한 네이버 뉴스 기사, 네이버 뉴스 댓글, 7개 방송사의 유튜브 뉴스의 댓글을 수집하여 인터넷 미디어 여론을 수집하였다. 이후 다양한 텍스트 방법론을 통하여 설문 텍스트와

인터넷 미디어 텍스트를 비교 분석하였다.

우선 제Ⅵ장에서는 설문 텍스트와 인터넷 미디어 텍스트들을 연관규칙 학습과 의미망 분석 방법론을 통하여 그 특징을 정성적으로 분석하였다. 연관규칙 학습에서 사용되는 주요 개념은 특정 단어 집합 A가 문장에 포함될 확률을 나타내는 지지도와 특정 단어 집합 B가 속한 문장에 다른 단어 집합 A가 포함될 확률인 신뢰도이다. 한편 의미망 분석의 경우 연관규칙 학습에서 얻어진 지지도가 높은 고빈도 단어를 중심으로 다른 단어들이 어떻게 연결되어 있는지 시각화를 통해서 텍스트의 구성을 분석한다.

분석 결과, 조세정책에 대한 설문 텍스트의 경우 고빈도 단어들의 지지도가 높아 사용되는 어휘가 정제된 것을 확인할 수 있었으며, 해당 조세정책에 대한 긍정 혹은 부정적인 선호에 대한 의견을 짐작할 수 있는 단어들을 신뢰도 및 의미망 분석을 통해 확인할 수 있었다. 네이버 뉴스 기사의 경우 설문과 같이 고빈도 단어들의 지지도가 높아 사용된 어휘가 정제되어 있었으나, 가상자산 기타소득세에 관한 긍·부정의 의견을 얻을 수 있을 만한 단어를 발견하기 어려웠다. 한편 네이버 뉴스 댓글이나 유튜브 뉴스 댓글의 경우, 사용되는 고빈도 단어들의 지지도가 상대적으로 낮아 어휘가 산발적이고 조세정책과 무관한 정치 관련 어휘들이 자주 등장하는 등 잡음이 많은 것을 확인할 수 있었다. 하지만 댓글에서 사용된 단어들은 해당 조세정책에 대한 긍·부정의 견해를 짐작할 만한 어휘들을 포함하고 있어 작성자의 의견을 어느 정도 유추할 수 있을 것으로 판단되었다.

이후 제Ⅶ장에서 잠재 디리클레 할당(Latent Dirichlet allocation, LDA)을 이용한 토픽모델링과 트랜스포머 모형을 사용한 감성분석을 통해 설문에서 얻어진 텍스트와 인터넷 미디어 여론을 정량적으로 분석하였다. LDA 토픽모델링은 텍스트를 유사한 분류로 묶어 주어 해당 텍스트들이 구성하고 있는 주제들의 구성비를 파악할 수 있게 한다. 한편 감성분석은 가상자산 기타소득세 신설에 대해 긍·부정의 견해를 명확하게 알 수 있는 설문의 텍스트를 트랜스포머 모형으로 학습시켜 감성분류기를 만든 후 인터넷 미디어 텍스트의 조세정책에 대한 긍·부정 비율을 분석해 보는 것이다.

LDA 토픽모델링을 하여 텍스트들을 구성하고 있는 주제들을 분류해 본

결과는 다음과 같다. 인터넷 미디어 여론들은 분석대상이 되는 가상자산 기타소득세에 관한 주제 이외에도 여러 주제들을 포함하고 있는 것을 확인할 수 있었다. 예를 들어 2018년의 네이버 뉴스 댓글 중 44%는 가상자산 기타소득세가 아닌 부동산에 대한 주제를 형성하고 있었다. 2019년 댓글도 34%는 전반적으로 증세 기조를 보이는 정부에 대한 불만을 주제로 하고 있었다. 2022년의 경우도 30%의 댓글은 금융투자소득세와 공매도에 관한 내용을 주제로 하고 있었으며, 27%의 경우에는 대선 후보들에 관한 내용을 주제로 하고 있었다.

반면 설문을 통해서 얻은 텍스트를 토픽모델링을 통해 분석하였을 때는 전반적으로 조세정책과 밀접한 관련이 있는 주제를 형성하고 있는 것을 확인할 수 있었다. 가상자산 기타소득세 신설과 관련하여 긍정으로 응답한 텍스트들은 불로소득에 대한 과세(14%), 수익에 대한 과세(20%) 등 응답자들이 생각하고 있는 조세정책에 대한 원칙들을 주제로 하는 경우가 많았으며, 부정적으로 응답한 텍스트들의 경우 투자자 보호(12%)와 과도한 세율(13%), 위험 감수에 따른 정당한 수익(20%) 등을 주제로 형성하고 있었다.

감성분석을 위해서는 트랜스포머 모형을 이용하였으며 설문 텍스트를 학습자료로 사용하였다. 학습된 감성분류기는 정확도, 정밀도, 재현율, F-1 score에서 0.86~0.93 수준의 높은 성능을 나타내었다. 설문에서는 가상자산 기타소득세에 대한 긍정 비율이 69.3% 정도를 보였기 때문에, 인터넷 여론이 국민들의 선호를 잘 반영한다면 유사한 긍정 비율이 나올 것으로 기대할 수 있다.

하지만 네이버 뉴스 댓글의 경우 약 8%만이 긍정으로 분류되었으며, 90% 이상이 부정으로 분류되었다. 분류기가 댓글에 대해서 잘못 분류할 오류의 가능성을 고려하더라도 긍정 분류는 16% 정도였으며 부정 분류는 80% 이상으로 나타나, 인터넷 미디어 여론은 조세정책에 대한 부정적인 여론을 매우 편향적으로 나타내고 있음을 보여주고 있다. 유튜브 댓글도 긍정 분류가 최대 27%, 부정 분류는 73%가량으로 예측되어 긍정 비율이 네이버 뉴스 댓글 보다는 더 높게 나타나기는 하였으나, 전반적으로는 설문 응답에 비해 부정적인 편향성이 있는 것을 확인할 수 있었다.

이처럼 설문과 인터넷 미디어 여론이 크게 다를 수 있는 이유에 대해서 추가로 검토해 보았다. 네이버 뉴스 댓글의 경우 댓글에 대해서 사용자들이 ‘좋아요’, ‘싫어요’라는 견해를 표출할 수 있다. 이러한 ‘좋아요 - 싫어요’ 수로 가중치를 부여하여 가상자산 기타소득세 신설에 관한 긍·부정의 분류를 수행하면 긍정 분류가 4%가 되고 부정 분류가 96%가 되어, 가중치가 없는 경우 (8% vs 90%)에 비해서 그 차이가 더욱 커졌다. 이는 긍정적으로 쓴 댓글보다 부정적으로 쓴 댓글이 훨씬 많은 ‘좋아요’를 받았다는 것인데, 이는 인터넷 여론을 작성하는 이들 대부분이 조세정책에 부정적인 의견을 가진 사람이었을 것임을 추측할 수 있게 한다.

이상의 결과를 종합하자면 특정 조세정책에 대한 인터넷 미디어 여론은 텍스트의 주제가 매우 산발적이어서 조세정책에 대한 선호를 잘 대표하지 못할 것으로 판단할 수 있다. 또한 감성분석 결과는 인터넷 미디어 여론이 조세정책에 대한 부정적인 의견을 매우 편향되게 표출하여 조세정책에 대한 국민들의 선호를 왜곡하여 반영할 가능성이 있음을 정량적으로 보여주고 있다. 이는 인터넷 미디어 여론을 의식하여 정책을 수정하거나 철회하는 것은 사회적 합의를 반복하는 것으로 공정하지 못하며 세계 안정성을 해칠 수도 있다는 주장을 뒷받침하는 결과이다.

본 연구는 특정 조세정책을 중심으로 수행된 기초연구로서 연구 결과를 일반화하기에는 추가적인 분석이 요구된다는 한계점이 있다. 하지만 가상자산 기타소득세 신설의 경우, 2020년 정부발의 세법개정안에서 시행 시기를 ‘2021년 10월 1일’에서 ‘2022년 1월 1일’로 변경하여 통과시켰으며, 2021년 의원발의 세법개정안으로 2023년 1월 시행하는 것으로 유예되었고, 2022년 정부발의 세제개편안에서는 2025년으로 유예하는 것을 논의하고 있는 상황으로, 본 연구의 결과와 밀접한 관련이 있을 것으로 생각한다. 직접적인 시사점으로는 인터넷 미디어 여론은 해당 조세정책에 대한 국민적인 선호를 잘 반영하지 못하고 편향되어 있기 때문에, 이를 정책 수립과정에 여과 없이 반영하는 것을 경계해야 할 것이라는 메시지를 제공한다.

좀 더 장기적이고 일반적인 시사점으로는 조세정책 입안 과정에서의 충분한 사회적 합의와 세계 안정성과 예측성을 높일 수 있는 방안이 필요함을 제시

한다. 이를 위해서는 경제학, 정치학 등 다양한 방면에서의 후속 연구가 필수적일 것이다. 구체적으로 국내 미디어 여론의 편향성에 관한 체계적인 연구, 정치적 양극화 및 여론 양극화의 측정 및 그 기제에 관한 연구, 조세정책에 관한 국민들의 선호를 도출하고 통합할 수 있는 체제에 관한 연구 등 여러 분야에서 본 연구보다 더 넓은 범위에서 발전된 방법론으로 기초연구, 정책연구, 학술적 연구가 지속적으로 수행되기를 기대한다.

목 차

I. 서론	17
II. 연구 배경 및 선행연구	23
1. 조세정책의 변화에 따른 연구 환경의 변화	23
2. 여론의 형성과 조세정책 의사결정과의 연관성 논의	31
3. 선행연구	39
가. 텍스트 분석 방법론(텍스트 마이닝)을 이용한 조세정책 연구	39
나. 여론의 편향성에 관한 연구	42
다. 납세인식에 관한 연구	44
III. 텍스트를 이용한 연구 방법론 개관	59
1. 텍스트를 이용한 연구 방법론	59
2. 본 연구에서 사용할 텍스트 분석 방법론	65
IV. 설문과 인터넷 미디어 여론 텍스트의 정성적 비교	69
1. 설문을 통해 수집된 조세정책 관련 여론 응답의 타당성 검토	70
2. 텍스트 플랫폼별 조세정책 관련 여론 분석	77
가. 텍스트 마이닝 방법론	77
나. 유튜브 댓글 분석	80
다. 가상자산 과세 관련 네이버 기사 및 댓글 비교 분석	86
라. 가상자산 과세 관련 설문조사 응답 분석 및 댓글과의 비교 분석	90
V. 인터넷 미디어 여론의 토픽 모델링 및 감성분석	97
1. 토픽 모델링 분석	97

가. 뉴스 기사와 댓글의 토픽 모델링	100
나. 조세정책 관련 설문 의 토픽 모델링	103
2. 조세 수용성을 측정할 감성사전의 구축	106
가. 감성사전의 정의와 구축 방법	106
나. 조세 수용성을 측정할 수 있는 감성사전의 구축	107
3. 감성사전을 활용한 인터넷 여론의 감성 측정	118
가. 뉴스 댓글을 활용한 감성사전의 검증	119
나. 댓글의 감성분류와 비교 분석	121
4. 설문조사 응답과 네이버 댓글 감성별 텍스트 심층 분석	123
VI. 결론 요약 및 시사점	129
참고문헌	133
부록	145
1. 국회예산정책처 보고서의 세법개정안 쟁점 모음	145
2. 설문조사 예시	162
3. 세종 품사 태그	169
4. 주제별 워드 클라우드	170

표목차

〈표 II-1〉 국회예산정책처 보고서의 세법개정안 쟁점 요약	24
〈표 II-2〉 연도별 세법개정안 관련 상위 키워드 및 빈도	33
〈표 III-1〉 문서를 수량화된 배열로 나타내는 방법들	62
〈표 IV-1〉 선행연구에서 주로 사용되는 변수에 따른 긍정 비율	71
〈표 IV-2〉 사회·인구통계학적 변수에 따른 긍정 비율	73
〈표 IV-3〉 긍정 응답에 대한 회귀분석 결과	74
〈표 IV-4〉 긍정 응답에 대한 회귀분석 결과(사회·인구통계학적 변수 포함)	75
〈표 IV-5〉 가상자산 과세 관련 유튜브 뉴스 댓글의 고빈도 단어 집합	81
〈표 IV-6〉 가상자산 과세 관련 유튜브 뉴스 댓글의 연관규칙	83
〈표 IV-7〉 가상자산 과세 관련 유튜브 뉴스 댓글의 중심성 분석	85
〈표 IV-8〉 연도별 분석 대상 문장 수	86
〈표 IV-9〉 가상자산 과세 관련 네이버 뉴스 기사 및 댓글 고빈도 단어 집합	87
〈표 IV-10〉 가상자산 과세 관련 네이버 뉴스 기사 및 댓글의 연관규칙	88
〈표 IV-11〉 가상자산 과세 관련 설문 응답의 고빈도 단어 집합	91
〈표 IV-12〉 가상자산 과세 관련 설문 응답의 연관규칙	92
〈표 IV-13〉 2022년 네이버 댓글과 설문 응답의 크기 1인 고빈도 단어 집합	94
〈표 V-1〉 암호화폐 과세 관련 기사 및 댓글 자료 건수	98
〈표 V-2〉 가상자산소득 과세 관련 기사 토픽 모델 분석 결과	101
〈표 V-3〉 가상자산소득 과세 관련 댓글 토픽 모델 분석 결과	103
〈표 V-4〉 조세정책 관련 개방형 설문에 따른 긍정·부정 응답 토픽 분석	105
〈표 V-5〉 ‘긍정’에서 제외한 답변	114
〈표 V-6〉 ‘부정’에서 제외한 답변	115

CONTENTS

〈표 V-7〉 표본별 크기와 분류 성능 평가지표	116
〈표 V-8〉 뉴스 댓글 분류 혼동행렬	119
〈표 V-9〉 부정으로 분류된 긍정 댓글	120
〈표 V-10〉 연도별 댓글의 감성	123
〈표 V-11〉 감성별 2022년 네이버 댓글과 설문 응답의 크기 1인 고빈도 단어 집합	124

그림목차

[그림 IV-1] 가상자산 과세 관련 유튜브 뉴스 댓글의 의미망 시각화	84
[그림 IV-2] 가상자산 과세 관련 네이버 뉴스 기사의 의미망 시각화	89
[그림 IV-3] 가상자산 과세 관련 네이버 뉴스 기사 댓글의 의미망 시각화	90
[그림 IV-4] 가상자산 과세 관련 설문 응답의 의미망 시각화	93
[그림 IV-5] 가상자산 과세 관련 2022년 네이버 뉴스 기사 댓글의 의미망 시각화 ..	95
[그림 V-1] 트랜스포머 모형 구조	111
[그림 V-2] 혼동행렬(Confusion Matrix)	115
[그림 V-3] 표본 설정에 따른 혼동행렬	117
[그림 V-4] 가상자산 과세 관련 설문 긍정 응답의 의미망 시각화	126
[그림 V-5] 가상자산 과세 관련 2022년 네이버 기사 긍정 댓글의 의미망 시각화 ..	127
[그림 V-6] 가상자산 과세 관련 설문 부정 응답의 의미망 시각화	127
[그림 V-7] 가상자산 과세 관련 2022년 네이버 기사 부정 댓글의 의미망 시각화 ..	128

I. 서론

“세상에 좋은 세금이란 없다”라는 윈스턴 처칠의 말처럼, 세금은 국가재정에 꼭 필요하지만 국민들에게는 영 달갑지 않은 부담일 것이다(박지웅 외 2018). 그렇기에 세금 부담이 증가하는 세법 개정이나 신규 세목의 신설은 국민적인 반대와 저항을 가져오기 쉬우며, 이러한 저항이 언론의 관심과 여론을 통해 형성되는 경우 정부가 본래 계획한 세율인상이나 공제혜택 축소와 같은 정책을 수정 혹은 폐지하는 경우도 발생하게 된다. 대표적인 사례는 2014년 귀속 근로소득세 연말정산 대란으로 2013년에 8월 8일 발표된 세법개정안은 근로소득자 세부담 증가에 대한 부정적인 여론이 생기자 곧이어 8월 13일에 수정안을 만들어 발표하였다(국회예산정책처, 2013). 그리고 2015년 초 2014년 귀속 연말정산을 받은 납세자들의 불만이 증대¹⁾되자 정부는 「2015년 연말정산 보완대책」을 발표하여 세액공제 혜택을 일부 확대하고 소급 적용하여 환급해 주었다(기획재정부 보도자료, 2015. 4. 6.).

이와 같은 국민들의 조세저항으로 인한 조세정책의 수정은 2020년 정부 제출 세법개정안에서 2021년 10월 시행하려고 했던 가상자산소득 기타소득세의 시행이 국회 심사 과정에서 2022년 1월 시행으로 변경되었다가, 2021년 의원발의 세법개정안으로 2023년 1월 시행으로 유예된 사례에서도 발견할 수 있다.²⁾ 또한 소액주주의 주식 양도소득에도 과세하는 금융투자소득 과세 체계를 도입했던 2020년 정부의 세법개정안도 기본공제액이 본래 논의되었던 2천만원보다 대폭 증가하여 5천만원으로 반영되어 제출된 사례가 있었다.³⁾

1) 근로소득자의 불만이 증폭된 일부 원인은 2012년 9월 간이세액표 개정으로 근로소득 원천징수액이 인하여 연말정산 기간에 환급금액이 적어진 것도 있다(김재진, 2016).

2) 2022 세계개발안에서 유예 논의 중이다(기획재정부, 『2022년 세계개발안』, 2022).

이상에서 살펴본 사례들과 같이 국민의 여론을 고려한 조세정책의 수정과정은 대체로 1년 내외의 짧은 기간에 이루어졌다.⁴⁾ 이렇게 상대적으로 짧은 기간에 정책적 의사결정을 하게 되는 경우, 해당 정책에 대한 충분한 검토나 국민적 의사 수렴과정이 배제되어 정책 방향의 과도한 선회가 발생할 수 있고,⁵⁾ 정책목표 달성보다는 인기에 영합하는 의사결정을 할 수 있으므로 정책 입안자들은 신중한 의사결정을 할 필요가 있다. 하지만 조세저항에 대한 언론의 관심과 여론을 어떻게 해석하고 어떻게 대응하는 것이 최적의 의사결정인지는 잘 알려져 있지 않다.

최근에는 텍스트 마이닝이라는 일컫는 방법론들이 개발되고 고도화됨에 따라 조세정책의 의사결정에 대해서 언론의 관심과 여론이 어떻게 반응하는지 분석하여 정부와 납세자 간의 상호작용이 정책 결정에 미치는 영향을 분석하는 시도가 이루어지고 있다. 황의찬(2021), 황의찬·우석진(2022)은 세금부담이 증가하는 조세정책 변화에 대해서 뉴스와 댓글을 수집하여 감성점수를 매긴 후, 감성점수가 정부가 계획 혹은 발표한 조세정책의 수정이나 철회에 미치는 영향을 분석한 바 있다. 해당 연구는 세금부담이 증가하는 조세정책과 관련된 인터넷 뉴스의 댓글을 모아서 긍·부정의 감성지수를 산출하였으며, 감성지수가 낮아질수록 정부가 정책을 철회할 가능성이 커진다는 것을 주장하였다.

그렇다면 이처럼 인터넷 뉴스 댓글과 같은 미디어에서 표출되는 여론이 조세정책의 수정이나 철회에 영향을 주는 것이 과연 사회적으로 바람직한 현상인지 의문을 제기할 수 있다. 정부의 조세정책의 변화로 인해서 납세의 부담을 더 지게 될 것으로 예상하는 납세자들이 인터넷 여론을 통해 자신들의 선호를 표출하고, 이것이 정부의 정책으로 반영되는 것은 사회적 합의를 만들어 가는 과정이라고 생각할 수 있다. 하지만 반대로 국민의 의사를 대표

3) 2022 세계개편안에서 유예 논의 중이다(기획재정부, 『2022년 세계개편안』, 2022).

4) <부록 1> 참고.

5) 예를 들어 근로소득자의 면세자 비율은 2013년에는 약 32.4%였으나, 2013년의 세법개정안의 수정발표와 2015년의 보완대책을 거치면서 2005년 수준(48.7%)인 48%로 급격히 증가하였다(김재진, 2016).

하는 국회에서 여야의 합의를 통해서 이미 국민적 합의에 이른 조세정책을 납세자들의 반발을 의식하여 수정 혹은 철회하는 것은 사회적 합의를 반복하는 것으로 공경하지 못하며 세제안정성을 해친다고 주장하는 의견도 존재한다(『경향신문』, 2021. 11. 3.). 결국 인터넷 미디어 여론이 조세정책에 영향을 주는 것이 바람직한가 하는 질문은, 해당 여론이 조세정책에 대한 국민의 선호를 얼마나 잘 반영하고 있는가 하는 질문으로 환원할 수 있을 것이다.

따라서 본 연구에서는 Gentzkow et al.(2019)의 이론적 틀을 바탕으로, 인터넷 미디어 여론이 조세정책에 관한 국민들의 선호를 얼마 잘 반영하는지 정성적·정량적으로 분석하고 이를 바탕으로 조세정책에 대한 시사점을 도출해보고자 한다.

본고의 구성은 다음과 같다. 제Ⅱ장에서는 연구의 배경과 선행연구를 소개하고자 한다. 우선 제Ⅱ장 제1절에서는 2013년부터 2021년까지 세법개정안의 주요 쟁점 사항을 정리한 후, 최근 몇 년간 세금부담이 증가하는 조세정책의 변화가 많아 이러한 정책변화에 따른 국민들의 인식을 연구하기에 좋은 환경이 조성되었음을 보이고자 한다. 제Ⅱ장 제2절에서는 2013년부터 2021년까지의 세법개정안 관련 인터넷 뉴스들의 단어의 빈도를 분석하여 뉴스들이 세법개정안의 쟁점들을 잘 반영하고 있는지, 여론의 형성에서 어떠한 특징을 가졌는지 간략히 검토한다. 제Ⅱ장 제3절에서는 선행연구들을 검토하고 본 연구와의 관련성을 논의한다.

제Ⅲ장에서는 텍스트를 이용한 연구 방법론(텍스트 마이닝)에 대해서 개괄적으로 소개한다. 이후 본 연구의 핵심 연구 질문인 인터넷 미디어 여론이 조세정책에 관한 국민들의 선호를 얼마나 잘 반영하는지를 연구하기 위해 활용한 텍스트 분석 방법론에 대해서 개괄적으로 소개한다. 이를 간략히 정리하면 다음과 같다. 우선 조세정책에 관한 국민 대표 여론을 설문을 통해 수집한다. 한편 조세정책에 관한 인터넷 미디어 여론은 네이버 뉴스 기사, 네이버 뉴스 댓글, 유튜브 댓글을 통해 수집한다. 이후 수집된 텍스트를 토큰화하여 수량화된 배열로 나타낸 후 정성적·정량적 분석을 통해 설문 텍스트와 인터넷 미디어 여론의 텍스트를 비교 분석한다.

제Ⅳ장은 가상자산 기타소득세 도입과 관련한 설문 텍스트와 인터넷 미디어 여론 텍스트의 정성적 비교 분석을 제시한다. 우선 설문을 통해 얻은 텍스트가 조세정책에 대한 국민들의 선호를 잘 반영하고 있는지를 확인할 필요가 있다. 하지만 국민들의 선호는 관찰 가능한 대상이 아니므로 이를 직접 확인하는 것은 불가능하다. 따라서 이러한 작업은 간접적으로만 수행될 수 있는데, 제Ⅳ장 제1절에서는 설문들의 응답들이 선행연구들에서 얻은 조세정책에 대한 선호들이 예측하는 응답과 유사한지를 검토하여 설문 응답의 타당성을 검토하였다(Sanity check).

제Ⅳ장 제2절에서는 수집된 설문 텍스트와 인터넷 미디어 여론 텍스트들을 대상으로 전처리를 수행한 후 토큰화(tokenize) 하여 수량화하였으며, 수량화된 텍스트들은 연관규칙학습과 의미망 분석을 통해 정성적으로 비교 분석하였다. 정성적인 분석 결과 인터넷 뉴스의 경우에는 가상자산 기타소득세 신설 정책에 대해서 한정된 어휘를 한정된 방식으로 사용하는 경향을 보였으며, 조세정책에 대한 긍·부정의 의견을 얻기는 어려운 것으로 보였다. 반면 인터넷 미디어 댓글의 경우 조세정책과 무관한 어휘들이 많이 발견되고 잡음이 많지만, 이를 적절히 제거하면 작성자들의 의견을 어느 정도 추출할 수 있는 것으로 나타났다. 예를 들어 가상자산소득 기타소득세 관련 유튜브 댓글과 네이버 댓글에서 투자자 보호, 가상화폐 금융자산 인정 의견과 관련된 단어들이 일관되게 나타나 텍스트가 댓글 작성자들의 생각을 잘 반영하고 있음을 시사한다. 설문 응답의 경우 조세정책에 관련성이 높은 정제된 언어로 이루어져 있음을 알 수 있었으며, 가상자산 기타소득세 신설에 대한 긍·부정의 의견을 파악하기 쉬운 것으로 나타났다.

이후 제Ⅴ장에서는 정량적인 분석을 위해 다음과 같은 작업을 수행하였다. 우선 가상자산 기타소득세 도입 관련 인터넷 미디어 여론의 텍스트를 잠재 디리클레 할당(Latent Dirichlet Allocation, LDA)을 통해 여러 주제로 분류한 후, 조세정책 주제와 밀접하게 관련된 텍스트의 비중이 얼마나 되는지 분석하였다. 분석 결과 인터넷 미디어에서 얻어지는 텍스트들은 조세정책과 관련된 주제도 일부 있었으나, 검색 키워드와 무관한 다른 조세정책에

관한 의견이라든가, 현 정부에 대한 일반적인 불만과 같이 조세정책과 관련이 없는 텍스트들도 많이 발견되었다. 예를 들어 가상자산소득 기타소득세 관련 네이버 뉴스 댓글에 대한 LDA 토픽모델링 분석 결과를 보면 2018년도의 댓글 44%가 부동산 관련 주제였으며, 2019년도의 댓글 중 35%는 금융투자소득세, 34%는 현 정부에 대한 일반적인 불만 등을 나타내는 주제로 파악이 되었다. 반면 가상자산 기타소득세 관련 설문조사에서 얻어진 텍스트들을 LDA 토픽모델링 분석을 했을 때는 일관성 있게 해당 조세정책에 대한 긍·부정의 선호를 잘 반영하고 있는 것으로 나타났다.

제V장 제2절에서는 설문 텍스트를 통해 조세정책에 대한 긍·부정의 감성을 도출할 수 있는 분류기를 만들어 인터넷 미디어 여론의 감성을 추정해 보고 이것이 어느 정도 편향적인지를 검토해 보았다. 설문 텍스트를 트랜스포머 모형을 통해 학습하여 감성분류기를 만들었으며, 구축된 감성분류기의 감성분류 성능은 정확도, 정밀도, 재현율, F-1 score가 모두 0.86~0.93 사이의 값을 가져, 매우 높은 것을 확인할 수 있었다. 이렇게 구축된 감성분류기를 통해 인터넷 미디어 여론 텍스트들을 감성분류해 본 결과, 부정 댓글은 95.7%를 부정으로 분류하였으나 긍정 댓글은 48.3%만 긍정 댓글로 분류하여 긍정 댓글에 대한 정밀도가 부족한 것을 확인할 수 있었다. 이는 댓글들의 경우, 사용된 단어가 과격한 경향이 있어서 조세정책에 대해서 긍정적인 의견을 표출할 때도 부정 감성으로 도출되는 경향이 있기 때문이었다. 이러한 오류의 가능성을 최대한 보수적으로 보정하여 인터넷 미디어 여론에 대해서 감성분류를 한 결과, 인터넷 미디어 여론은 가상자산 기타소득세 신설에 대해 매우 부정적으로 편향된 것을 확인할 수 있었다. 설문의 경우 가상자산 기타소득세에 대한 긍정 비율이 69.3% 정도였으나, 네이버 뉴스 댓글의 경우 긍정 비율이 최대 16%, 부정 비율은 최소 80% 이상으로 분석되었다. 유튜브 댓글도 긍정 비율이 최대 27%, 부정 비율은 최소 73%로 나타났다. 한편 네이버 댓글에서 독자들이 표시하는 ‘좋아요’와 ‘싫어요’ 정보를 이용하여 ‘좋아요’를 많이 받은 댓글에 더 큰 가중치를 부여하여 감성분석을 한 결과, 부정의 강도가 더욱 높아지는 것을 확인할 수 있었다. 이는 해당 조세

정책에 대해서 긍정적으로 쓴 댓글보다 부정적으로 쓴 댓글이 훨씬 많은 ‘좋아요’를 받았다는 것을 의미하며, 관련 조세정책에 대한 인터넷 여론을 조성하는 이들 대부분이 해당 조세정책에 부정적인 의견을 가진 사람이었을 것임을 추측할 수 있게 한다.

제VI장은 연구의 내용을 요약하고 시사점을 제시한다. 본 연구의 결과는 인터넷 미디어 여론이 조세정책에 관한 국민들의 선호를 잘 반영하기보다는 세부담이 증가하는 조세정책의 도입에 반대하는 의견을 편향되게 나타낸다는 것을 보여주고 있다. 이러한 점을 기초해 본다면 황의찬·우석진(2022)에서 보여준 것과 같이 조세정책이 인터넷 미디어 여론에 영향을 받는 것은 바람직하지 않은 현상이라고 할 수 있다. 특히 정치적 성향이 양극단에 있는 사람들이 온라인 여론 형성에 적극 참여한다는 최근의 연구 결과를 고려해 본다면(임원혁 외, 2019), 조세정책과 같이 정치적 여론 수렴과 사회적 합의가 중요한 정부 정책은 인터넷 미디어 여론의 영향을 받기보다는 국민적 선호가 잘 수렴될 수 있도록 하는 제도와 환경을 구축하는 것이 중요하다고 할 수 있겠다.

II. 연구 배경 및 선행연구

본 장에서는 먼저 최근 2013년부터 2021년까지의 세법개정안을 검토하여 조세정책 변화에 대해서 정리한 후, 최근 몇 년간 세금부담이 증가하는 조세정책의 변화가 많아 이러한 정책변화에 대한 국민들의 인식을 연구하기 좋은 연구 환경이 조성되었음을 보이고자 한다.⁶⁾ 둘째로는 2013년부터 2021년의 세법개정안 관련 뉴스들의 단어의 빈도를 분석하여 뉴스들이 세법개정안의 쟁점들을 잘 반영하고 있는지 검토한다. 셋째로는 인터넷 미디어 여론이 정부의 조세정책 결정에 미치는 영향을 분석한 황의찬(2021), 황의찬·우석진(2022)의 연구를 비롯한 선행연구들을 소개한다. 이를 통해 텍스트 마이닝 기법을 이용해서 여론을 수치화하여 조세정책 변화에 미치는 영향을 분석하는 것에 유용성과 그 한계에 대해서 논해 보고자 한다. 이러한 배경에 대해서 이미 익숙한 독자들은 본 장을 생략하고 제Ⅲ장으로 넘어가는 것을 추천한다.

1. 조세정책의 변화에 따른 연구 환경의 변화

조세법률주의에 따라 조세정책을 바꾸기 위해서는 세법이나 시행령을 개정하여야 한다. 정부는 일반적으로 연 1회 세법개정안을 발표하고 국회에 세법개정안을 제출한다. 하지만 매년 세부담이 증가하는 세법개정안이 제출되는 것은 아니기 때문에 세부담 증가에 따른 국민들의 여론 형성이나 이에 따른 정책철회 같은 현상은 정기적으로 관찰되는 것은 아니다. 하지만 최근

6) 이러한 인식은 다른 연구들에서도 비슷하게 발견된다. 황의찬·우석진(2022)의 경우 2013~2020년 동안 12개의 세부담 증가에 따른 조세저항 이벤트를 포착하였는데, 2013~2016년도에는 4개가 포착되었으나 2017~2020년도에는 8개의 조세저항 이벤트가 포착되었다.

몇 년간은 국민의 삶에 밀접한 영향을 끼쳐 세금부담이 증가되는 세법개정 안들이 다수 있었으며, 이러한 상황은 세금부담 증가에 대한 국민들의 여론 형성과 이 여론들이 조세정책 결정 과정에 미치는 영향을 분석하기 좋은 환경을 조성하고 있다.

〈표 II-1〉 국회예산정책처 보고서의 세법개정안 쟁점 요약

구분	2013년	특이사항
소득세제	소득세 공제방식의 변경	2013년 수정 반영 2015년 재수정
	근로장려세제 확대 및 자녀장려세제 신설	
	신용카드 공제율 인하	미반영
	종교인 과세	2014년 유예
법인세제	성장동력 확충을 위한 세제지원 확대	
	일자리 창출을 위한 세제지원	
소비세제	치료 이외 미용·성형용역의 면세대상 제외	
재산과세	일감몰아주기 증여세 과세제도 완화	
	상증세법상 기업승계 적용대상 확대 등	
구분	2014년	특이사항
소득세제	배당소득 증대세제	
	비과세종합저축 등 과세특례 저축상품 통합 재설계	
법인세제	기업소득 환류세제	
	근로소득 증대세제	
	고용창출투자 세액공제	
소비세제	금융·보험용역의 부가가치세 면세범위 축소	
	중고자동차에 대한 매입세액공제 축소	
재산과세	담배가격 인상에 따른 세제 개편	통과
	「상속세 및 증여세법」상 기업승계 적용대상 확대 등	
구분	2015년	특이사항
소득세제	비과세 개인종합자산관리계좌 도입	
	상장주식 양도소득세 대주주 범위 확대 등	
	종교소득 과세	2018년 시행
법인세제	청년고용증대세제	
	R&D 부문 세액공제	
소비세제	법인 등 업무용 승용차 관련 비용 과세 합리화	
	개별소비세 과세대상 및 기준가격, 세율 조정	

〈표 II-1〉의 계속

구분	2016년	특이사항
소득세제	신용카드 소득공제 적용기한 연장	
	주택 임대소득 세제지원 비과세 특례 연장	
	상장주식 과세대상 대주주 범위 확대	
법인세제	신성장동력·원천기술 R&D 세액공제 확대	
	영상콘텐츠 제작비용 세액공제 신설	
소비세제	재활용폐자원 등 의제매입세액공제 적용기한 연장	
재산과세	가업상속공제 제도 합리화	
구분	2017년	특이사항
소득세제	고소득층 과세 강화	통과
	대주주 주식양도차익 과세 강화 (3억원 초과 25% 과세, 대주주 범위 2021년 3억원)	2020년 대주주 범위 확대 철회
	다주택자 양도소득세 인상 등	2019~2020년, 2022년 한시 유예
	근로장려금 지급 확대 산정액 상향 조정	
	장기보유특별공제 합리화	
법인세제	법인세율 인상	통과
	고용증대세제(청년고용증대세제)	
	대기업 연구인력개발비 세액공제 축소	
소비세제	신용카드사 부가가치세 대리납부제도 신설	
	궤련형 전자담배 개별소비세 인상	통과
재산과세	일감몰아주기 과세 강화	
	가업상속공제 공제요건 강화	
	상속·증여 신고세액공제 축소	
지방세제	궤련형 전자담배 세율인상에 따른 담배소비세율 조정	통과
구분	2018년	특이사항
소득세제	근로장려세제(EITC) 확대	
	주택 임대소득 과세 개편	
	신용카드 등 소득공제 일몰 연장	
	기부금 세액공제 확대	
법인세제	위기지역 지원세제 신설	
소비세제	부가가치세 납부면제점 인상	
	노후경유자동차 교체에 대한 개별소비세 감면 신설	
재산과세	종합부동산세: 세율 인상	2021년 일부 완화
	공정시장가액비율 조정	

〈표 II-1〉의 계속

구분	2018년	특이사항
재산과세	종합부동산세: 세부담 상한비율 인상	
	장기보유 1세대 1주택에 대한 세액공제율 인상	
	종합부동산세: 분납대상 확대 및 분납기간 확대	
	상속세 및 증여세: 기업상속공제의 사후관리요건 완화 등	
구분	2019년	특이사항
소득세제	근로소득공제 한도 신설	
	신용카드 등 사용금액에 대한 소득공제 적용기한 연장 등	
법인세제	접대비 손금산입 한도 상향	
	내국법인의 소재·부품·장비 전문 기업 출자·인수에 대한 과세특례	
소비세제	탁주와 맥주의 과세체계 전환	
	수소전기차 개별소비세 감면 적용기한 연장	
	노후차 교체 시 개별소비세 감면	
	증권거래세를 인하	
재산과세	기업상속공제 제도의 사후관리요건 완화	
	최대주주 할증평가 제도 완화	
구분	2020년	특이사항
소득세제	금융투자소득 과세체계 도입	공제금액 수정 통과 2022년 유예
	소득세 최고세율 인상	통과
	가상자산소득 기타소득세 신설	2020년 유예 2021년 재유예 2022년 재유예
	장기일반민간임대주택 등에 대한 양도소득세 과세특례 축소	
법인세제	통합투자세액공제 재설계	
	기업환경 개선을 위한 세제개편	
	특정 내국법인의 초과유보 소득 배당간주 신설	미반영
소비세제	간이과세 기준금액 인상 등	
	납부면제 기준금액 인상	
	과세대상 담배의 확대	
	역상형 전자담배에 대한 개별소비세율 인상	미반영
재산과세	증권거래세를 인하	
	부동산 세제개편(양도소득세, 법인세 포함)	의원안 발의 2021년 일부 완화

〈표 II-1〉의 계속

구분	2021년	특이사항
소득세제	근로장려금 소득기준금액 인상	
	1세대 1주택 양도소득세 비과세 기준금액 상향	의원안 발의
	가상자산소득 과세시행 1년 유예	의원안 발의
	기부금세액공제 한시 상향	
법인세제	국가전략기술 R&D·시설투자 공제율 신설 등 투자세액공제 확대	
재산과세	기업상속공제 대상기업 범위 확대	
	영농상속공제 한도 상향 등	
	상속세 물납 대상 확대	
	1세대 1주택자에 대한 주택분 종합부동산세 기본공제액 상향	정기국회 이전 통과 (의원안)~

주: 세부담이 증가하는 것이 명백한 쟁점의 경우에는 수정 철회 없이 통과된 경우 통과를 명시함
 자료: 국회예산정책처, 『2013~2017년 세법개정안 분석』, 『2018~2020년 시행 개정세법의 주요 내용 및 심사 쟁점』, 『2020~2021 개정세법의 심사 경과와 주요 내용』을 참고하여 저자 작성

〈표 II-1〉은 국회예산정책처(National Assembly Budget Office, NABO)에서 매년 발간하는 세법개정안 분석과 심사 경과 및 쟁점을 정리한 보고서에서 소득세제, 법인세제, 소비세제, 재산과세 관련 쟁점들을 모아둔 것이다.^{7) 8)} 〈표 II-1〉을 통해서 말하고 싶은 것은 2013년 소득세 공제방식을 소득공제에서 세액공제로 전환하고 2014년에 담배세를 인상한 이후, 세금부담을 크게 변화시킬 만한 변화가 별로 없다가 2017년 들어서 세금부담이 증가되는 세법개정안과 세목이 다수 신설되었다는 것이다. 이러한 변화는 세금부담 증가에 따른 여론의 변화, 그리고 그 여론의 변화에 따른 정책 결정자

- 7) 소비세 분야는 「소득세법」, 「조세특례제한법」 소득세 분야를 포함하며, 법인세 분야는 「법인세법」, 「조세특례제한법」 법인세 분야, 소비과세는 「부가가치세법」, 「개별소비세법」, 「조세특례제한법」 소비과세 분야 등을, 재산과세 분야는 「상속세 및 증여세법」, 「종합부동산세법」, 「증권거래세법」, 「조세특례제한법」 재산과세 분야를 포함한다(국회예산정책처, 2021).
- 8) 가독성을 높이기 위하여 쟁점들 중 국민들이 체감하는 세부담의 변화와 관련이 있을 만한 쟁점들을 중심으로 발췌하였으며, 전체 쟁점들은 〈부록 1〉에 소개하였다. 법인세제의 경우 「조세특례제한법」의 일몰연장 여부 관련 쟁점이 다수였으나, 단순한 일몰연장의 경우 조세정책의 기간만 바뀌는 것이고 세금부담의 변화가 없으므로 제외하였다. 부가가치세의 지방소비세 이양비율 인상과 같은 쟁점들도 국민들이 체감하는 세금부담과는 연관이 없으므로 제외하였다.

들의 대응과 이에 따른 조세정책의 변화에 대해 연구할 수 있는 환경을 제공하고 있다고 할 수 있다. 시간순으로 세금부담에 변화를 준 세법개정안의 쟁점을 중심으로 간략히 설명하면서 세금부담을 가져다주는 세법개정이 원래의 의도대로 개정되었는지, 아니면 중도에 철회되거나 유예 혹은 수정되지 않았는지 부가적으로 설명하고자 한다.

2013년에 개정된 「소득세법」 개정안은 소득 공제방식을 소득공제에서 세액공제로 전환하는 것을 골자로 하고 있다. 하지만 세법개정안이 발표되자마자 세금부담이 커지는 근로자들의 반발이 있었고, 개정안이 발표된 지 5일 만에 수정되었다(국회예산정책처, 2013).⁹⁾ 이후 개정된 세법이 적용되는 2014년 귀속 근로소득 연말정산 기간에는 과거보다 연말정산을 적게 받게 된 근로자들의 반발이 있었고 정부는 「2015년 연말정산 보완대책」을 발표하여 세액공제 혜택을 일부 확대하고 소급 적용하여 환급하는 일이 있었다. 신용카드 소득공제율 축소와 관련된 개정안의 경우, 소득공제율을 15%에서 10%로 하향 조정하려고 하였으나 국회 본회의에서 통과되지 못하였다. 종교인 과세의 경우 2013년 「소득세법 시행령」 개정을 통해 종교인의 소득에 대해서 2015년부터 기타소득으로 과세하고자 하였으나 2014년 개정안에서 1년 유예되었으며, 2015년 개정안에서 「소득세법」 개정을 통해 2018년부터 시행하게 되었다.

한편 2014년 정부는 경제장관회의를 통해 흡연으로 인한 사회적 피해를 줄이기 위해 담배가격 인상을 포함한 금연종합대책을 추진기로 하였다(국회예산정책처, 2014). 담배가격을 2,500원에서 4,500원으로 인상하고, 이를 위해 담배소비세, 지방교육세, 국민건강증진부담금을 인상하고 개별소비세를 추가하였다. 해당 개정안은 검토과정에서 개별소비세가 종가세에서 종량세로 수정되기는 하였으나, 최종적으로 국회 본회의에 상정되어 가결되었다(황의찬·우석진, 2022).

이후 2015년과 2016년에는 세금부담의 큰 변화를 줄 만한 세법개정이 없

9) 2013년에 8월 8일 발표된 세법개정안은 근로소득자 세부담 증가에 대한 부정적인 여론이 생기자 곧이어 8월 13일에 수정되었다.

었다. 국회예산정책처(2015)의 세법개정안 평가에 의하면 2015년 세법개정안은 중장기 세입 확충 방안에 대해서 보다 적극적인 검토가 필요하다고 평가하였으며, 2016년의 세법개정안에 대해서는 주목할 만한 개정안 내용이 없이 대부분 기존 제도의 적용 범위를 조정하고 일몰 기간을 연장하는 등의 소극적인 개정에 그치고 있다고 평가하였다(국회예산정책처, 2015; 2016).

문재인 정부가 들어선 2017년은 세금부담을 증가시킬 만한 많은 세법개정이 있었다. 소득세 분야에서는 최고세율 인상 및 3억원 초과~5억원 이하 과표구간 신설 및 세율 인상이 있었으며, 대주주 주식 양도소득세율 인상과 대주주 범위 확대가 있었다. 또한 다주택자에 대한 양도소득세 증과가 있었다. 이 중 대주주 범위 확대의 경우 2021년 4월부터 3억원 이상 보유자로 강화될 예정이었으나, 2020년 시행령 개정안에서 10억원 이상 보유자로 요건이 유지되었다. 한편 다주택자 양도세 증과의 경우 2019년 12월 「주택시장 안정화 방안」(기획재정부 보도자료, 2019. 12. 16.)에 따라 다주택자가 조정대상지역에서 10년 이상 보유한 주택을 양도하는 경우 한시적으로 증과를 배제해 주었으나, 2020년에는 ‘7·10 대책’을 발표(국토교통부, 2020)하여 다주택자 양도세 증과를 더욱 강하게 하였고, 2022년 5월 10일 기준으로 양도세 증과가 한시 배제(기획재정부 보도자료, 2022. 5. 9.)되는 등 많은 개정이 있었다.¹⁰⁾ 한편 법인세제의 경우 최고세율 구간이 신설되고 세율이 인상되었으며, 결련형 전자담배 개별소비세가 인상되었고, 이에 따라 지방교육세가 인상되었다. 해당 개정안은 중대한 수정 없이 통과되었다.

2018년 종합부동산세를 강화하는 세법개정안이 발표되었다. 정부의 세법개정안은 7월 30일에 발표되었지만 세법개정안 발표 이후로도 서울을 중심으로 부동산 가격이 급등하자 과표구간 신설, 세율인상, 공정시가액비율 상향조정, 세부담 상한비율 상향 등을 포함하는 ‘9·13 대책’을 발표(기획재정부 보도자료, 2018. 9. 13.)하였고, 이후 부동산시장이 안정되며 ‘9·13 대책’ 관련 의원안을 포함한 대안이 의결되었다(국회예산정책처, 2018). 종합부동산세의

10) 이러한 변화는 납세자들의 저항뿐만 아니라 정부가 조세정책을 통해 부동산시장의 경기 조절 수단으로 사용했기 때문이라고 볼 수 있다.

강화는 부동산시장의 안정화를 위한 목적을 가지고 있었으나, 소득이 적고 집을 팔기 어려운 1세대 1주택자들에게도 과도한 부담을 준다는 평가를 받기도 하였다.¹¹⁾ 이후 2021년에는 1세대 1주택자들의 주택분 종합부동산세 기본공제액이 상향되는 법안이 의원안으로 발의되어 정기국회 이전에 통과되어 1세대 1주택자들의 부담이 완화되었다.¹²⁾

2019년의 경우 일몰 예정이었던 신용카드 소득공제 혜택이 축소될 수도 있다는 논란이 있었으나, 기획재정부에서 보도자료를 배포하여 신용카드 소득공제 연장을 전제로 검토하고 있다고 발표하였다(황의찬, 2021; 기획재정부 보도참고자료, 2019. 3. 11.). 한편 수입 주류와 국산 주류의 과세형평성 확보를 위해 주세가 종량세로 전환되고 증권거래세가 일부 인하되는 등의 변화가 있었으나 세금부담에 큰 영향을 줄 만한 변화는 없었다.

2020년에는 소득세 최고세율이 인상되었으며 새로운 세목이 신설되었는데, 금융투자소득 과세체계가 도입되어 대주주가 아닌 개인 투자자들도 주식양도차익에 대한 세금을 부담하게 되었으며, 가상자산소득에 대해서는 기타소득세가 신설되었다. 소득세 최고세율 인상안은 통과되었으며 금융투자소득 과세체계를 2023년부터 시행하는 개정안은 기본공제액이 본래 논의되었던 2천만원보다 증가한 5천만원으로 개정안에 반영되어 통과하였다. 가상자산소득에 대해 과세하는 개정안은 시행 시기를 2021년 10월에서 2022년 1월 시행으로 수정·통과되었다. 또한 이는 2021년 의원발의 개정안을 통해 1년이 추가로 유예되어 2023년부터 시행하는 것으로 개정되었다.¹³⁾ 한편 2020년 세법개정안은 다주택자 및 단기보유 거래에 대한 과세를 강화하고, 법인의 주택 양도소득 과세를 강화하였으며, 법인·다주택자 등의 주택 보유과세를 강화하는 방향을 포함하고 있었는데, 이는 정기국회 이전 의원안

11) 『이데일리』, 「집 Talk」 억울한 1주택자...“우리도 투기꾼이냐”, 2020. 1. 24., <https://n.news.naver.com/mnews/article/018/0004561670?sid=101>, 검색일자: 2022. 10. 2.

12) 이와 유사한 법안은 2019년도에도 의원안으로 발의되어 검토된 바 있다(국회예산정책처, 2019, p. 132).

13) 금융투자소득 과세체계와 가상자산소득 기타소득세의 경우 2022년 세제개편안에서 유예 검토 중이다.

으로 발의되어 8월 4일 통과하였다. 정부의 개정안에 포함되었던 개인 유사 법인의 초과 유보소득 배당간주 과세 신설안과 전자담배 개별소비세 인상안은 미반영되었다.

2021년도의 경우에는 세금부담을 일부 완화하는 개정안이 주로 의원안을 통해서 이루어졌다는 특징이 있다. 소득세의 경우 1세대 1주택 양도소득세 비과세 기준 금액이 상향되었으며, 가상자산소득 과세는 1년 유예되었다. 종합부동산세도 1세대 1주택자에 대한 주택분 종합부동산세 기본공제액 상향이 정기국회 이전 의원안으로 통과되었다.

이상을 살펴보면 2014년 이후 큰 세법개정안이 없다가 2017년부터 세금 부담이 증가하는 세법개정안이 다수 있었음을 확인할 수 있다. 세금부담 증가를 가져오는 개정안은 수정 없이 통과되는 예도 있으나, 미반영되거나 수정·통과되거나 혹은 통과되었더라도 후에 시행이 유예되거나 개정되는 경우가 있는 것도 확인할 수 있었다. 이는 세금부담이 완화되는 개정안은 대부분 수정 없이 통과되는 것과 대조적이며, 이는 납세자들의 저항이나 반발이 조세정책에 영향을 미쳤을 가능성을 시사한다. 특히 2021년의 경우 의원안 발의를 통해서 이전 세법개정을 통해서 증가한 세금부담을 낮추는 개정안을 통과시킨 예도 있었는데, 이는 세금부담에 반발하는 국민들의 여론을 반영하는 움직임으로 해석할 수 있는 여지가 있다.

2. 여론의 형성과 조세정책 의사결정과의 연관성 논의

이상에서 확인할 수 있듯이 세부담이 증가하는 세법개정안의 경우 국회에서의 의결과정에서 수정 혹은 철회되거나 또는 이후에 세부담을 완화하는 방안으로 개정되는 경우가 자주 발생하는 것을 알 수 있다. 그렇다면 이러한 현상들은 세부담을 느낀 납세자들의 여론 형성으로부터 영향을 받은 것이라고 할 수 있을까? 황의찬·우석진(2022)은 납세자의 여론이 정책 결정 과정에 영향을 줄 것이라는 사회적 인지도도 불구하고 조세저항이 정책 결정에 미치는 영향을 수치화한 연구는 발견하기 어렵다고 언급하였다. 이는 납세자들의 여론 형성은 인터넷 뉴스 댓글이라든가, 청와대 국민청원, 소셜

미디어 등과 같이 텍스트의 형태로 나타나기 때문에 수치화하기 어려운 성격이 있기 때문이라고 할 수 있다.

하지만 최근 발달하고 있는 텍스트 마이닝 기법은 뉴스나 댓글과 같은 텍스트 정보를 수량화된 배열로 나타내어 분석하는 이론적 틀을 제시하여 텍스트를 정량적으로 분석할 수 있는 기반을 제공하고 있다. 본 절에서는 간단한 단어 빈도 분석 방법을 통해 인터넷 뉴스들이 세법개정안의 쟁점들을 잘 반영하고 있는지 검토해 본다.

〈표 II-2〉는 2013년부터 2021년까지 네이버 뉴스에서 ‘세법개정안’이라는 키워드를 검색했을 때 나오는 뉴스 제목을 모아서 각 연도의 상위 35개 단어의 빈도를 측정해 본 것이다.¹⁴⁾ 이는 텍스트마이닝 방법론 중 간단한 키워드 빈도 분석 방법이지만 조세정책에 대한 여론의 관심을 짐작해 볼 수 있게 한다.

이하에서는 각 연도의 상위 키워드들이 국회예산정책처의 보고서에 쟁점과 연결이 잘 되는지 특히 세부담이 증가하는 쟁점에 대해서 밀접하게 연관이 되어있는지 확인해 보고자 한다. 만일 상위 키워드들이 국회예산정책처 보고서들의 쟁점과 잘 연결이 된다면 언론이 주목하는 쟁점들과 전문가들이 생각하는 쟁점들이 어느 정도 일치한다는 것을 의미할 것이고, 이는 뉴스를 분석하는 것을 조세정책에 대한 분석 방법의 하나로 사용할 수 있는 정당성을 부여해 줄 것이다. 한편 뉴스가 세부담이 커지는 조세정책 이슈에 대해서 더욱 민감하게 반응한다면, 세부담에 대한 조세저항이나 수용성을 연구하는 데 뉴스를 이용하는 것에 대한 유용성을 제공할 것이다. 반면 뉴스가 세부담이 커지는 조세정책에 민감하게 반응한다는 것은 뉴스가 편향성도 가지고 있다는 것을 의미하므로, 뉴스를 중립적인 미디어로 보기는 어렵다는 시사점도 제공할 것이다.

14) 세법개정안이라고 검색했을 때 자연스럽게 따라서 나오기 쉬운 단어들, 예를 들어 ‘국회’, ‘여야’, ‘기재부’ 등은 제외하였으며, 대통령이나 기재부 장관과 같은 사람의 이름도 제외하였다. 또한 ‘축소’, ‘확대’, ‘감면’, ‘증가’, ‘기준’, ‘폐지’와 같이 다른 단어와 함께 해석하지 않고는 어떠한 의미를 찾기 어려운 단어들도 제외하였다.

〈표 II-2〉 연도별 세법개정안 관련 상위 키워드 및 빈도

순위	2013년	2014년	2015년	2016년	2017년	2018년	2019년	2020년	2021년	빈도
1	부담	289 소득공제	223 연말정산	1249 법인세	429 증세	597 중부세	424 세액공제	263 부동산	468 양도세	320
2	증세	268 사내유보금	167 종교인	237 세액공제	229 법인세	415 양도세	211 양도세	402 세액공제	402 세액공제	193
3	중산층	262 종교인	150 세금폭탄	236 소득공제	186 일자리	285 부동산	191 투자	123 중부세	382 반도세	192
4	세부담	214 증세	139 업무용	203 신용카드	123 부자증세	231 보유세	189 주택	113 주식	273 중부세	183
5	소득공제	185 신용카드	129 법인세	198 증세	119 대기업	189 다주택자	154 맥주	110 소득세	206 가산자산	179
6	재검토	168 투자	101 폭탄	147 소득세	77 부동산	170 일자리	142 기업상승공제	108 주택	194 상속세	149
7	종교인	160 면세한도	94 소금적용	147 카드	75 소득세	170 개편	136 전환	103 코로나	171 주택	128
8	세액공제	136 법인세	92 환급	138 세제지원	75 종교인	160 대책	118 소득공제	96 다주택자	163 투자	126
9	신용카드	126 직장인	90 보완책	138 투자	73 부자	115 외국인	116 증권거래세	94 증세	159 법인세	114
10	연봉	121 대기업	89 대기업	134 대기업	72 소득공제	99 자녀장려금	114 증량세	89 증권거래세	150 부동산	111
11	원점	113 부가세	84 민생통장	104 공익법인	63 세액공제	99 임대사업자	110 소득세	88 대책	137 주택자	109
12	소득세	110 경제활성화	76 개별소비세	100 양도세	60 부담	96 부담	106 개편	87 세액공제	135 코로나	104
13	근로자	100 부담	74 증세	96 월세	58 전자담배	88 등록	98 소재	85 주택자	122 주식	98
14	수정	99 가계소득	74 보완대책	95 고소득자	57 대책	86 세액공제	95 부품	84 가산자산	122 비트코인	85
15	복지	98 세액공제	70 부가세	95 부담	53 최고세율	82 주택자	85 최저임금	82 대주주	117 대기업	84
16	대기업	88 배당소득	69 부담	93 면세자	52 확정	79 대주주	72 대기업	79 최고세율	115 0벤트	83
17	서민	88 세습유대자축	68 세액공제	91 둘째	52 고소득자	78 최고	71 세제지원	77 부가세	114 증개형	81
18	체크카드	87 우려	62 대책	82 세제혜택	51 서민	76 부과	71 제로페이	72 가산화폐	112 감세	81
19	소득	83 유예	59 세부담	78 부가세	51 신설	73 맥주	71 개소세	70 투자	110 배타리	77

〈표 II-2〉의 계속

순위	2013년	2014년	2015년	2016년	2017년	2018년	2019년	2020년	2021년	빈도
20	법인세	83 연말정산	56 청년	78 출산	50 소득	73 집값	71 상속세	69 간이과세	105 시행	빈도
21	즉시연금	78 배당	56 근로자	77 면세점	48 전쟁	71 개선	71 기업활력법	64 중소기업	97 유예	빈도
22	세금포탄	72 서민	55 소득세법	73 법인세율	48 투자	68 개편안	67 부담	63 후속	96 미술품	빈도
23	소득공제율	70 퇴직연금	55 무늬만	70 중소기업	47 양도세	68 투자	64 규제	63 금융세제	88 백신	빈도
24	공약기계부	66 세부담	52 소금	67 전쟁	47 보유세	65 법인세	63 감세	62 부담	87 가격	빈도
25	직장인	66 증대세제	50 면제	67 임대소득	44 고용	64 방안	62 신용카드	61 임호화폐	86 분양권	빈도
26	수장안	66 부자감세	48 양도세	63 경제활력	42 부가세	64 혁신성장	62 중부세	61 유보소득세	84 근로장려금	빈도
27	물이주기	60 체크카드	48 차량	62 일자리	41 우려	64 증세	59 제외	60 유보소득	81 보유세	빈도
28	정비	59 기업소득환류세제	46 직장인	59 부자	40 아이코스	64 우려	59 면제	58 부과	79 가상화폐	빈도
29	월급쟁이	59 양도세	45 구글세	59 이하	39 담뱃세	64 신용카드	59 부동산	58 소득공제	77 면제	빈도
30	국민	58 부동산	45 이하	57 원상복합	37 입법전쟁	63 증광세	58 법인세	58 전자담배	74 청년	빈도
31	없는	57 기업승계	44 남세지연맹	57 소득	37 한도	62 주택	57 한도	58 소득	73 기부금	빈도
32	임금0주기	57 담배값	44 평균	55 서민	37 초고소득자	62 저소득층	57 세부담	58 폭탄	72 쿠팡리커질	빈도
33	일감	57 제외	43 추가환급	55 법인	35 증시	61 소득공제	56 수출규제	56 거래세	71 부담	빈도
34	양도세	56 담뱃세	43 우려	53 한도	35 결련형	61 이하	56 주세	55 신설	71 일부터	빈도
35	질세	54 카드	41 수익	53 의장	35 창출	60 부가세	56 부과	55 한도	70 경총	빈도

자료: 저자 작성

국회예산정책처(2013)가 제시한 소득세 쟁점(〈표 II-1〉 참고)은 소득세 공제방식 변경, 근로장려세제 확대, 신용카드 공제율 인하, 종교인 과세 4가지이다. 2013년도 상위 키워드를 보면 ‘소득공제(5)’, ‘종교인(7)’, ‘세액공제(8)’, ‘신용카드(9)’가 있으므로 세금부담이 완화되는 근로장려세제를 제외한 3개의 쟁점을 잘 반영하고 있다. 한편 키워드 ‘몰아주기(27)’, ‘일감몰아주기(32)’를 보면 국회예산정책처(2013)가 제시한 일감몰아주기 증여세 과세제도 완화 쟁점을 반영하고 있는 것으로 보인다. 이 외에 특징적인 것은 상위 키워드 중에서 ‘재검토’, ‘원점’, ‘수정’, ‘수정안’ 등은 세법 개정안이 중도에 수정된 것을 반영하고 있고, ‘부담’, ‘근로자’, ‘세금폭탄’, ‘월급쟁이’ 등도 상위 키워드에 등재되어 소득세 공제방식 변경으로 인해 세부담이 커지는 근로자들에 대한 인식을 보여주고 있다.

2014년도는 정부가 가계소득 증대를 통한 경기활성화를 위한 세제를 집중적으로 발표한 시기이다. 국회예산정책처(2014)가 제시한 쟁점은 배당소득 증대세제, 기업소득 환류세제, 근로소득 증대세제 등이 있었으며, 2014년도의 상위 키워드인 사내유보금(2), 경기활성화(12), 가계소득(14), 배당소득(16), 배당(21) 등은 이러한 쟁점을 반영하고 있다. 한편 국회예산정책처(2014)는 소비세제에서 담배가격 인상에 따른 세제개편을 제시했는데, 담뱃값(32), 담뱃세(34)¹⁵⁾의 키워드를 통해 이러한 쟁점을 반영하고 있다. 한편 종교인(3) 키워드는 동 보고서가 제시한 쟁점이 아닌데도 불구하고 높은 순위를 차지하였는데, 이는 2013년도 개정안에 포함되었던 종교인 과세가 2014년도에 1년 유예되었기 때문이다.

2015년도의 상위 키워드를 보면 ‘연말정산(1)’이라는 키워드가 1249회 등장하여 모든 연도를 통틀어서 압도적으로 많이 등장하는데, 이는 2013년도 세법개정으로 인해 2014년 귀속 근로소득세에 대한 연말정산 금액이 크게 바뀌었기 때문이다. 상위 키워드를 보면 세금폭탄(3), 폭탄(6) 등이 있어서 연말정산 때에 세금부담이 커진 근로자들의 상황을 반영하고 있다. 당시 연

15) 두 키워드의 경우 각각 32번째, 34번째로 자주 등장하지만 두 키워드의 빈도와 〈표 II-2〉에는 나타나지 않은 ‘담배가격’ 등의 키워드와 합치면 상위 7번째 키워드가 된다.

말정산으로 인해 여론이 악화되자 정부는 환급해 주기로 하였고, 이것이 소급적용(7), 환급(8), 보완책(9), 보완대책(14) 등의 키워드에 반영되어 있다. 국회예산정책처(2015)의 소득세 관련 쟁점이었던 비과세 개인종합자산관리계좌 도입은 만능통장(11)이라는 키워드, 종교소득 과세 쟁점¹⁶⁾은 종교인(2)이라는 키워드에 반영되었다. 다른 상위 키워드로 업무용(4), 차량(27)이 있으며, 이는 동 보고서에 포함된 업무용 승용차 관련 비용 과세 합리화 쟁점과 관련이 있다. 또한 개별소비세(12) 키워드가 2015년에 유일하게 상위 키워드에 포함되었는데, 이는 보고서상의 쟁점인 개별소비세 과세대상 조정¹⁷⁾과 연관된다.

2016년의 경우 국회예산정책처(2016)가 주목할 만한 것이 없는 개정안이라고 평가하였을 만큼 쟁점이 없는 해였다. 이를 반영하듯이 100회 이상 나오는 키워드가 5개 밖에 없는데, 이는 모든 연도를 통틀어 가장 적은 숫자이다. 소득세제 쟁점인 신용카드 소득공제 적용기한 연장의 경우 신용카드(4) 키워드에 반영되어 있으며, 주택임대소득 세제지원 비과세 특례 연장은 임대소득(25)과 연관이 있다. 국회예산정책처(2016)에 쟁점이 많지 않아 이외의 쟁점은 키워드에서 찾아보기 어렵다. 특이한 점은 법인세(1)라는 키워드가 많이 언급되었다는 것인데, 이는 정부의 세법개정안에는 포함되어있지 않았지만 당시 야당이었던 민주당, 국민의당, 정의당에서 법인세 인상안을 의원안으로 발의했기 때문이다. 이는 뉴스 텍스트가 세법개정안에는 반영되지만 당시 이슈가 되는 쟁점들을 잘 포함하고 있음을 보여준다.

2017년도에는 문재인 정부가 취임하면서 세금부담을 증대시키는 변화가 많은 세법개정안이 발의되었다. 국회예산정책처(2017)의 소득세제 관련 쟁점은 고소득층 과세 강화, 대주주 주식양도차익 과세 강화, 다주택자 양도소득세 인상, 근로장려금 지급 확대, 장기보유특별공제 합리화가 있었으며, 2017년의 상위 키워드인 부자증세(4), 부동산(6), 최고세율(15), 고소득자(17),

16) 2013년 개정안 시행령에 반영되었다가 2014년에 1년 유예된 이후 2015년 개정안에 법률로 반영하고, 세법개정 후속 시행령 개정을 통해 2018년부터 시행되었다.

17) 녹용, 로열젤리, 향수를 과세대상에서 제외하고, 보석·귀금속, 시계 등 일부 과세물품의 기준가격을 200만원에서 500만원으로 상향하였다(국회예산정책처, 2015).

양도세(23), 초고소득자(32) 키워드는 고소득층 증세와 다주택자 양도소득세 쟁점과 관련이 있다. 한편 법인세제 관련해서는 대기업(5), 최고세율(15) 키워드들이 법인세율 인상, 대기업 연구 인력개발비 세액공제 축소 쟁점과 연관된다. 소비세제 쟁점과 관련하여서는 궤련형 전자담배 개소세 인상이 있었으며, 이는 전자담배(13), 아이코스(28), 담뱃세(29), 궤련형(34)과 관련된다. 2017년도에 특징적인 키워드는 증세(1)라는 키워드가 597회로 가장 많이 등장했으며, 소득세, 법인세의 과표구간 신설로 인해 신설(19)이라는 키워드가 등장하였다.

2018년도는 종합부동산세가 대폭 개정된 해였으며 국회예산정책처(2018)도 부동산 관련 쟁점을 주로 언급하였다. 상위 키워드 5개인 종부세(1), 양도세(2), 부동산(3), 보유세(4), 다주택자(5)가 부동산 관련 키워드들이었으며, 임대사업자(11), 주택자(15), 집값(20) 또한 부동산 세제와 관련이 있다. 한편 신용카드 소득공제 일몰연장 쟁점과 관련하여 신용카드(29)가 언급되었다. 한편 세법개정안에는 포함되지 않았으나 맥주를 종량세로 개편하는 논의가 있었기에 맥주(19)라는 키워드가 등장했다.

2019년도에는 큰 세제 변화가 없었던 해였으며, 이에 따라 키워드의 빈도도 다른 해에 비해서는 상대적으로 낮은 것을 확인할 수 있다. 국회예산정책처(2019)의 소비과세 쟁점인 탁주와 맥주의 과세체계 전환은 맥주(5), 종량세(10), 주세(34)와 관련이 있으며, 증권거래세율 인하의 경우 증권거래세(9) 키워드와 관련이 있다. 법인세제의 경우 소재·부품·장비 기업에 대한 과세특례를 신설하는 쟁점이 있었으며, 이는 소재(13), 부품(14) 키워드에 반영되어 있다. 소득세 관련 쟁점인 신용카드 소득공제 적용기한 연장과 관련하여서는 신용카드(25)가 언급되었다.

2020년도에는 세금부담이 증가하는 세법개정이 많았던 해이다. 국회예산정책처(2020)는 소득세제의 쟁점으로 금융투자소득 과세체계 도입, 소득세 최고세율 인상, 가장자산소득 기타소득세 신설, 장기일반민간임대주택 등에 대한 양도소득세 과세특례 축소를 꼽았으며, 이들 쟁점은 양도세(2) 주식(4), 가상자산(14), 최고세율(16), 가상화폐(18), 암호화폐(25)와 같은 키워드와 밀

접하게 연관된다. 한편 키워드 중 유보소득세(26), 유보소득(27)은 법인세제 쟁점인 특정내국법인의 초과유보 소득 배당간주¹⁸⁾와 관련이 있다. 소비세제 관련 쟁점에서는 간이과세 기준금액 인상은 간이과세(20) 키워드와 관련이 있으며, 액상형 전자담배에 대한 개소세 인상¹⁹⁾과 과세대상 담배 확대 쟁점은 전자담배(30)에 반영되어 있다. 한편 2020년에는 정부안이 아닌 의원안으로 발의된 부동산 세제가 개정되었는데, 이러한 쟁점과 관련해서는 부동산(1), 종부세(3), 다주택자(8), 주택자(13)와 연관이 있다.

2021년도의 경우에는 세금부담을 일부 완화하는 개정안이 의원안을 통해서 이루어졌다는 특징이 있다. 국회예산정책처(2021)는 의원안으로 개정된 쟁점들 중 1세대 1주택 양도소득세 비과세 기준금액 상향과 가상자산소득 과세시행 1년 유예, 1세대 1주택 종부세 기본공제 상향을 언급하였으며, 이는 양도세(1), 종부세(4), 가상자산(5), 주택(7), 비트코인(14), 가상화폐(28)와 관련되어 있다.²⁰⁾ 정부안으로 개정된 쟁점은 소득세 관련 근로장려금 소득 기준금액 인상, 기부금 세액공제 한시 상향이 있었고, 이는 근로장려금(26), 기부금(31) 키워드와 직접적인 연관이 있다. 법인세 쟁점과 관련해서는 국가전략기술 R&D·시설투자 공제가 신설되었는데, 이는 국가전략기술(32) 키워드에 반영되어 있고, 재산과세 관련 쟁점인 상속세 물납 대상 확대 쟁점의 경우에는 미술품(22)과 관련되어 있다.

이상의 키워드 분석을 보면 뉴스의 상위 키워드들은 연도별로 크게 변화하며 매년 국회예산정책처에서 발간하는 세법개정안 분석보고서의 쟁점들과 밀접한 연관이 있는 것으로 보인다. 또한 쟁점이 별로 없는 것으로 평가되는 세법개정안이 제출된 2016년에는 키워드의 빈도수도 눈에 띄게 줄어들어 키워드의 빈도도 세법개정안에 대한 정보를 포함하고 있는 것으로 보인다. 특히 뉴스의 상위 키워드들은 세금부담이 완화되는 세법개정안 쟁점보다는 불균형적으로 세금부담이 늘어나는 쟁점을 반영하는 경향을 보였다. 이는

18) 미통과.

19) 미통과.

20) 개정안이 주로 1세대 1주택자에 한정되었기에 이전 연도에 나왔던 다주택자 키워드는 나타나지 않은 것으로 보인다.

뉴스가 편향된 성질을 가지는 것을 보여주는 하지만, 세금부담이 증가되는 개정안에 대한 납세자들의 인식을 연구하는 데 유용한 정보원이 될 수도 있음을 시사한다고 할 수 있을 것이다.

3. 선행연구

가. 텍스트 분석 방법론(텍스트 마이닝)을 이용한 조세정책 연구

황의찬(2021), 우석진·황의찬(2022)은 텍스트 마이닝 방법론 중 감성분석(sentiment analysis) 혹은 오피니언 마이닝(opinion mining)이라고 불리는 방법론을 사용하여 납세자의 정서가 정부의 조세정책에 미치는 영향을 분석하였다. 황의찬(2021)의 경우 2019년에 논의되었다가 철회된 소득공제 혜택 축소 정책과 2020년도에 논의되어 기본공제액이 2천만원에서 5천만원으로 변경되어 개정안에 반영된 금융투자소득 과세체계²¹⁾에 대한 감성분석을 실시하여 납세자들의 정서가 이 두 정책의 철회 또는 수정에 어떠한 영향을 미쳤는지 분석하였다.

이러한 분석을 위해서 우선 저자는 '신용카드 소득공제'라는 키워드로 네이버 뉴스와 댓글을 수집하여 단어를 어절로 나누고 용언과 체언만을 추출하여 출현빈도가 40회 이상 나오는 단어에 대해서 감성점수를 부여하였다. 이것을 조세 온톨로지(감성사전)라고 지칭하였으며, 이 조세 온톨로지에 범용적으로 사용되는 범용 온톨로지²²⁾를 결합하여 통합 온톨로지를 구축하였다. 그리고 이 온톨로지를 사용하여 신용카드 소득공제 혜택축소와 금융투자소득 과세체계에 대한 감성지수를 산출하였으며, 해당 감성지수와 정부 정책이 어떠한 연관성이 있는지 분석하였다.

한편 황의찬·우석진(2022)은 2013~2020년 사이에 '세법개정안'이라는 키워드로 뉴스를 검색하여 조세저항이 있었을 것으로 예상되는 12개의 개정안을

21) 2023년도부터 시행 예정이나 2022년 세제개편안에서 유예에 대한 논의가 이루어지고 있다.

22) 저자는 군산대학교에서 개발한 KNU 감성사전을 사용하였다.

택하고,²³⁾ 이 개정안에 대한 감성지수를 산출하였다.²⁴⁾ 그리고 이들 개정안이 수정 없이 통과되었을 경우와 수정 혹은 철회된 경우를 나누어 감성지수가 정책 수정 혹은 철회 확률에 어떠한 영향을 미쳤는지 분석하였다. 로지스틱 회귀분석을 수행한 결과, 납세자의 부정적인 정서를 대표하는 감성지수가 1단위 낮아지면 정부가 정책을 철회할 가능성이 약 1.5% 증가하는 것으로 나타났다.²⁵⁾ 한편 조세 감성지수로 정책을 수정·철회할 가능성을 예측한다면 조세 감성지수 -69.87을 기준으로 수정·철회를 예측하는 것이 가장 예측력이 높은 것으로 보였다.

한편 2017년에 국회예산정책처에서 발간한 『트렌드 세법-과거 20년간 세법개정안의 궤적을 담다』(국회예산정책처, 2017a)는 과거 20년간(1996~2016) 의원 및 정부가 발의 및 제출한 총 2,701건의 세법개정안을 분석대상으로 하여 워드 클라우드 기법, 관계망 분석 방법론 등을 통해 분석하였다. 2,701건의 법안 중 의원발의 법안은 2,410건이며 정부제출 법안은 291건으로 의원발의안이 90%가량을 차지하는데, 이는 민주화의 진전, 여소야대, 법안실명제 등의 영향으로 의원발의안의 수가 시간에 따라 급격히 늘어났기 때문이다(국회예산정책처, 2017a).²⁶⁾ 이처럼 의원발의 세법개정안이 급격히 늘어나는 환경에서는 다량의 텍스트를 효율적으로 분석할 수 있는 텍스트 마이닝 방법론이 유용하게 사용될 수 있을 것으로 보인다.

해당 연구에서 기간별로 단어 빈도를 분석²⁷⁾한 결과, 제19대 국회 이후 ‘형평’, ‘중소기업’, ‘고용’ 키워드가 많이 발견되었으며, 글로벌 금융위기 직후인 2010년부터 ‘양극화·격차’, ‘형평’ 등의 키워드 빈도수가 증가하기 시작하여 2016년에는 전년 대비 두 배에 가까운 증가세를 보였다. 증세 여부별로 나

23) 2013년 소득세 공제방식 변경, 2014년 담뱃세 신설, 2017년 소득세·법인세 최고세율 인상 등 12개의 조세저항 이벤트(우석진·황의찬, 2022).

24) 황의찬(2021)에서 구축된 감성사전을 사용하였다.

25) 통계적으로 유의성이 높지는 않았다.

26) 법안실명제의 영향으로 2000년대부터 의원발의안이 늘어나기 시작했으며, 2010년부터 이러한 현상이 가속화되고 있다(국회예산정책처, 2017a).

27) 키워드 분석은 법안의 제안 이유에 한정하여 이루어졌다.

누어 단어 분석을 했을 때는 증세법안은 주로 ‘형평’, ‘분배’, ‘양극화’라는 단어를 포함하여 주로 부동산 및 주식 등 양도소득세 법안 위주로 발의되었으며, ‘고령화’, ‘재정건전성’이라는 키워드도 발견할 수 있었다. 감세법안은 주로 ‘중소기업’, ‘농업’ 등 취약계층을 위한 세제지원이 모든 기간에 걸쳐 고루 나타나는 경향이 있었다. 한편 모든 연도의 키워드 분석에서 ‘부동산’, ‘양도’ 등의 단어 수의 빈도가 높게 나타나 조세정책을 활용하여 부동산시장을 관리하려는 경향이 있다는 것을 확인할 수 있었다.

또한 해당 연구는 키워드 분석 이외에 단어들 간의 관계망 분석을 하였는데, 이러한 분석 방법은 단어가 일정 범위 내에서 같이 등장하는 빈도를 파악하기에 유용한 방법이다. 전 연도에 걸친 관계망 분석 결과, ‘형평’, ‘중소기업’, ‘고용’은 다른 단어들과 동시에 언급되는 빈도가 높았으며, 이는 각각 세법개정안이 해당 키워드들을 주요 정책목표로 설정하고 이를 중심으로 정책 수단과 정책 타깃을 설정하였음을 추론하게 한다. 예를 들어 ‘형평’의 경우 ‘대기업’, ‘중소기업’, ‘자영업자’, ‘장애인’, ‘택시’, ‘외국인’, ‘농업’ 등과 같은 단어들과 같이 나타나는데, 이는 형평이라는 정책목표의 주요 타깃들을 의미하는 것으로 해석 가능하다고 본 연구는 주장하였다. 한편 ‘재정건전성’이라는 키워드는 ‘저출산’, ‘고령화’ 사회에 대비하여 ‘대기업’과 ‘고소득층’을 주로 타깃으로 하는 정책 수단을 사용하는 것으로 추론할 수 있는 관계망을 보여주었다.

본 연구는 텍스트 분석 방법론을 사용하여 조세정책 관련 인터넷 미디어 여론이 조세정책에 관한 국민들의 선호를 잘 반영하고 있는지 연구한다. 황의찬·우석진(2022)은 세부담이 증가하는 조세정책에 대한 부정적인 인터넷 미디어 여론이 정부가 해당 조세정책을 수정하거나 철회하는 데 영향을 주는 것을 보였지만, 인터넷 미디어 여론이 조세정책에 대한 국민들의 실제 선호를 잘 반영하고 있는지는 논하지 않았다. 텍스트 마이닝 방법론은 여론의 편향성을 연구하는 데 자주 활용되어 왔으므로 본 연구의 목적에 적합할 것으로 보인다. 다음 절에서는 텍스트 마이닝 방법론을 이용한 여론의 편향성 연구를 소개하도록 한다.

나. 여론의 편향성에 관한 연구

Groseclose and Milyo(2005)는 텍스트 마이닝 방법론을 통해 언론사들의 정치적 편향성을 정량적으로 연구할 수 있는 방법을 제시했다는 측면에서 그 기여도가 높다. 이들은 미국의 언론사들의 정치적인 성향을 측정하기 위하여 다음과 같은 방법론을 사용하였다. 미국 정치인들의 경우 그들의 투표 기록을 통해서 그들의 정치적 성향을 측정하는 Americans for Democratic Action(ADA) 점수가 있다. 이 점수를 통해서 정치인들의 정치적 성향을 측정 후, 이들 정치인들의 연설에서 사용되는 단어들이 각각의 언론사에서 얼마나 자주 활용되는가를 통해서 언론사들의 정치적 성향을 정량적으로 측정하였다. 간단히 말해 진보(보수)적인 정치인들이 자주 쓰는 단어들을 자주 사용하는 언론사는 진보(보수)적인 것으로 측정한다는 것이다.

구체적으로 Groseclose and Milyo(2005)는 정치인들이 사용하는 단어들 중 Think Tank와 정치단체의 이름에 주목하였다. 이러한 단체들의 경우에는 정치적인 성향이 분명히 존재하므로 정치인들의 정치 성향에 따라 이 이름들이 언급될 빈도가 크게 다르다. 따라서 이러한 단체들이 언급되는 빈도를 주목해서 본다면 언론사들이 진보적인 정치인이 자주 언급하는 언어를 자주 인용하는지 아니면 보수적인 정치인이 자주 거론하는 언어를 사용하는지 구분이 잘 된다는 것이다. 분석 결과 언론사들은 정치인들보다 중도적인 경향을 가지는 경우가 많았으며, 평균적으로는 약간 진보적인 경향이 있는 것으로 확인되었다.

Gentzkow and Shapiro(2010)의 경우 Groseclose and Milyo(2005)의 방법론을 일부 개선하여 미국의 신문사들의 정치적 성향을 측정하였다. 저자들은 연방의회 의사록(Congressional Record)에 사용된 단어들 중에서 발언자의 소속정당을 가장 잘 예측할 수 있는 단어들을 통계적으로 추출하여 Groseclose and Milyo(2005)가 사용한 Think Tank와 정치단체 이름을 대체하였다. 이러한 방법론은 정치적 성향을 측정하는 데 사용되는 단어들의 선택에 있어서 자의성을 많이 감소시키는 데 기여했다.

Zhao et al.(2011)의 경우 잠재 디리클레 할당(Latent Dirichlet Allocation,

LDA)이라는 방법론을 통해 트위터(twitter)와 전통적인 언론인 뉴욕타임스(The New York Times)가 다루는 뉴스의 주제들이 어떻게 다른지 분석하였다. 저자들은 LDA 모형을 통해 트위터와 뉴욕타임스가 다루는 뉴스들의 주제를 분류하였으며, 분류된 주제들은 사건중심(event-oriented), 인물중심(entity-oriented), 일반(long-standing) 카테고리 분류하였다. 분류 결과 트위터의 경우는 뉴욕타임스보다 인물중심의 뉴스가 상대적으로 더 많고 사건중심 뉴스의 빈도는 낮은 것으로 확인되었다.

김진주·윤종빈(2021)의 경우 제21대 총선 유튜브 개표방송의 라이브 채팅창의 텍스트를 분석하여 언론사별 여론이 양극화되어 있는지 분석하였다. 총선 개표방송 1시간 동안 TV조선, JTBC, KBS, 연합뉴스 실시간 채팅을 수집하였으며, 형태소 분석을 통해 다빈도어와 연관어를 분석한 뒤 채팅의 내용을 감정적·이념적으로 구분하여 내용을 분석하였다. 저자들의 경우 실시간 채팅의 내용을 세부적으로 진보·보수 성향으로 구분하여 점수를 매기고 언론사별로 살펴보았는데, 정치적 성향이 뚜렷한 것으로 알려져 있는 JTBC와 TV조선의 실시간 채팅에서 이념 성향이 편향적으로 나타나, 반향실 효과를 일으킬 수 있는 동질성 강화를 중심으로 이념이 표출되고 있다는 것을 확인하였다.

이새미 외(2020)의 경우 코로나19 확산 시기 '마스크 5부제'에 대한 온라인 뉴스 기사와 카페글 텍스트를 수집·분석하여 언론 매체별 차이점을 알아보았다. 단어 빈도분석, 워드 클라우드, LDA 토픽모델링 분석을 실시한 결과, 온라인 뉴스에 비해 온라인 카페의 경우 '대리 구매', '개학 연기', '마스크 사용', '마스크 구입'과 같이 실생활 관련 토픽이 나타나, 개인 미디어의 특성이 반영됨으로써 정보 전달의 기능보다는 개인의 의견, 감정, 정보를 교류하는 역할을 하는 것으로 판단하였다.

이상에서 텍스트 마이닝을 통한 여론의 편향성 연구에 관한 논의를 살펴보았다. 이러한 연구들은 방법론적으로 크게 두 가지의 형태를 가지고 있다. 첫 번째는 Groseclose and Milyo(2005), Gentzkow and Shapiro(2010), 김진주·윤종빈(2021)과 같이 정치적 성향에 대한 점수를 매길 방법을 마련한

다음, 이를 언론의 텍스트에 적용하여 언론의 정치적 성향을 측정하는 것이다. 두 번째는 Zhao et al.(2011)이나 이새미 외(2020)와 같이 텍스트들을 여러 주제로 분류한 다음, 이러한 주제들이 차지하는 비율이나 특징들을 분석하는 방법이 있을 수 있다. 본고에서는 이러한 방법들을 모두 활용하여 정량적인 분석을 시도한다(제V장의 내용).

한편 임원혁 외(2019)의 경우에는 텍스트 분석 방법을 사용하지는 않았으나 여론의 양극화 현상을 설문자료 분석과 실험 방법론을 통하여 분석하였다. 이들의 분석 결과는 정치적 성향의 분포상 양극단에 있는 사람들이 현실정치와 온라인의 여론 형성에 적극 참여하고 있으며, 이러한 현상이 더욱 강화되고 있다는 것을 보여주고 있다. 조세정책의 경우 사회적 합의가 중요한 만큼, 만일 인터넷 미디어 여론이 조세정책에 대한 국민의 선호를 반영하기보다는 정치적 성향의 극단에 있는 사람의 의견을 편향되게 반영한다면, 이는 조세정책의 수립과정에 인터넷 미디어 여론의 영향이 미치는 것은 바람직하지 않음을 주장하는 근거가 될 것이다.

다. 납세인식에 관한 연구

조세정책에 대한 납세자들의 선호는 여론을 형성하고 조세정책의 결정 과정에 영향을 줄 것이라는 사회적 인지에도 불구하고 여론의 형성과 조세정책 결정의 상호작용을 분석하는 연구는 많지 않다. 기존의 선행연구들은 주로 이미 존재하는 조세정책, 특히 소득세에 대한 납세자들의 납세순응 인식 연구를 통해 조세정책에 대한 납세자들의 선호를 연구하였다. 본 절에서는 이러한 기존의 선행연구들을 소개하고자 한다. 기존의 선행연구들은 텍스트 분석을 통해 수행할 본 연구의 핵심 주제와는 직접적인 연관성이 적다. 하지만 본 연구의 설문에서 수집된 데이터들이 신뢰할 만한 외적 타당성을 가졌는지 확인하는 과정에서 기존 선행연구에서 밝혀진 사실들을 사용할 예정이므로 이를 소개하고자 한다. 텍스트 분석을 통한 본 연구 핵심 결과에 주된 관심이 있는 독자들은 본 절을 생략하고 제Ⅲ장으로 넘어갈 것을 추천한다.

납세순응에 관한 경제학적인 접근을 한 연구는 Allingham and Sandmo (1972)로부터 시작했다고 할 수 있다. 해당 연구에서는 납세자들은 기대효용을 극대화하기를 원하는 합리적인 경제 주체라고 가정했으며, 소득의 크기, 세율, 탈세 적발확률, 처벌의 강도가 주어질 때 경제적인 이익을 최대화하는 소득신고 금액을 선택하는 문제로 납세순응의 문제를 접근하였다. 해당 모형에서는 소득의 크기나 세율의 경우에는 모형에 따라 납세순응에 미치는 영향이 불분명하였으나, 적발확률과 처벌의 크기에 따라서는 납세순응의 크기가 올라가는 것을 확인하였다. 이러한 이론적 접근 방식은 이후의 경제적 억지모형(Economic Deterrence model)으로 불리는 연구들이 발전하는 데 영향을 미치게 되었다.

Yitzhaki(1974)는 탈세적발 시의 벌금이 신고누락분이 아닌 탈루세액에 비례하는 모형을 제시하였다. 이러한 모형하에서는 세율의 증가가 탈세의 감소를 가져오는 결과를 얻는 것을 보였다. Srinivasan(1973)은 비례세와 누진세를 포함하는 일반적인 소득세함수를 가정하여 분석하였다. 납세자는 위험 중립적이며 일반적인 소득세함수를 가정했을 때, 해당 연구는 적발확률이 감소함에 따라 소득신고 수준도 감소한다는 결과를 얻었다. 또한 누진세제하에서 소득이 증가하면 과소소득신고 비율 역시 증가하지만, 비례세제하에서 적발확률이 소득의 증가함수인 경우에는 과소소득신고의 비율은 소득이 증가함에 따라 감소한다는 것을 보였다.

이러한 이론연구들의 예측은 소득의 크기, 세율, 적발확률, 처벌 강도의 변화가 납세순응에 미치는 영향을 분석하는 실증연구들로 이어진다. 이러한 연구들을 다 소개하는 것은 지면상 어렵지만 대표적인 연구들을 위주로 소개하고자 한다.²⁸⁾

1) 소득수준 변화에 따른 납세순응 변화

소득수준이 납세순응에 미치는 영향에 대해 크게 양(+의 관계와 음(-)의 관계로 나눌 수 있다. 먼저 Slemrod(1985)는 1977년 미국 재무부 세금 데

28) 좀 더 자세한 리뷰를 위해서는 Kirchner et al.(2010)을 참고할 수 있다.

이터를 이용하여 소득과 성실납부가 음(-)의 관계가 있음을 분석하였으며, Weck-Hannemann and Pommerehne(1989)의 스위스 납세자 데이터를 활용한 연구에서도 고소득자일수록 세금 납부를 회피하며 납세순응도가 낮다고 나타났다.

반면 소득과 납세순응의 양(+)의 관계를 분석한 실증연구로는 미국 국제청의 납세자 준수 측정 프로그램 데이터를 활용한 Christian(1994) 연구가 있다. 연간 소득이 10만달러 이상인 납세자는 실제 소득의 96.6%를 보고했지만 소득이 2만 5천달러 미만인 납세자는 실제 소득의 85.9%만을 보고하여 소득수준이 높을수록 납세순응도가 높게 나타났다. Fishlow and Friedman (1994)은 아르헨티나, 브라질, 칠레를 대상으로 실험한 결과, 낮은 소득수준에서 납세순응도가 감소하는 것으로 나타나 경제성장률이 낮고 인플레이션이 높은 국가에서도 소득과 납세순응도는 양(+)의 관계를 보인다고 하였다.

그러나 소득수준과 납세순응도 사이의 뚜렷한 관계가 나타나지 않은 연구들도 다수 있다. Park and Hyun(2003)의 한국을 대상으로 한 연구에서 납세자들은 소득수준과 관계없이 동일한 정도의 순응도를 가지고 있으며, 이는 CRRR 효율함수의 증거로 해석될 수 있다. 한편 여러 요인들은 소득수준과 납세순응의 관계를 모호하게 하는 역할을 한다. 예를 들어, 고소득자일수록 전문 세무사를 고용하여 탈세를 하지 않고도 법의 허점을 이용해서 절세를 할 수 있으므로 납세순응도가 높은 것처럼 보일 수 있다(Slemrod et al., 2001; Wärneryd and Walerud, 1982). 그리고 절대적인 소득보다는 고용의 형태가 납세순응에 미치는 영향이 클 수 있는데, 자영업자의 경우 고용된 납세자보다 탈세의 기회가 더 많을 수 있을 것이다.

2) 세율에 따른 변화

경제 이론에 따르면 세율이 높은 경우 실질소득이 감소하여 탈세의 수익성이 높아지나 절대위험 회피도도 증가하여 상쇄효과가 발생하므로 명확한 예측이 어렵다. 하지만 대부분의 실증연구에서는 세율이 높을수록 납세순응도가 더 낮은 것으로 나타났으며, 세율과 납세순응도 사이의 양(+)의 관계를

보이거나 효과가 없는 것으로 나타난 연구는 일부에 불과하다. 먼저 음(-)의 관계가 발견된 연구부터 살펴보면, Clotfelter(1983)와 Slemrod(1985)가 미국의 납세자 데이터와 재무부 세금 데이터를 활용하여 탈세가 한계세율에 민감하다는 계량경제학적인 증거를 제시하였으며, 미국뿐만 아니라 독일, 스위스 사례의 연구에서도 높은 한계세율에서 낮은 납세순응도가 나타났다(Lang et al., 1997; Weck-Hannemann and Pommerehne, 1989).

반대로 세율과 납세순응의 양(+)의 관계에 관해 연구한 논문은 많지 않으나 Feinstein(1991)의 미국 국세청 납세자 준수 프로그램 데이터를 활용한 연구에서 세율이 높을수록 납세순응도가 높은 것으로 나타났다. 대부분의 실증연구 결과들은 세율이 납세순응에 부정적인 영향을 미친다고 말하고 있으나, 세율의 변화는 소득을 변화시키기에 세율의 효과와 소득 효과를 분리해내기가 어렵다. Boylan and Sprinkle(2001)은 소득원에 따라 세율과 납세순응도의 관계가 달라지는지를 실험을 통해 연구하였는데, 실험 설계자로부터 소득을 부여받았을 경우 더 낮은 순응도로 세율 인상에 반응했으나 직접 번 소득에 대해서는 더 높은 순응도로 세율 인상에 반응하였다는 결론을 통해, 세율의 변화에 대한 납세자의 반응은 소득을 창출하는 데 필요한 시간과 노력의 양에 좌우될 수 있다고 주장하였다. 또한 Moser et al.(1995)은 세율 변화에 따른 납세자의 반응이 경제적 효과와 수평적 불평등 및 교환 불평등에 대한 인식에 의존하는지 실험을 통해 연구하였으며, 그 결과 높은 세율의 부정적인 영향은 다른 사람과 비교하여 불평등하게 대우받았다고 인식되는 경우에만 나타난다고 하였다. 이처럼 세율의 변화에 따른 납세순응도의 관계는 소득원, 인지된 세율의 공정성 등이 영향을 끼칠 수 있다.

3) 적발확률에 따른 변화

경제 모델은 납세자가 성공적인 회피의 이득과 적발 및 처벌 위협을 비교하여 납세 준수 결정을 한다고 가정한다. 탈세할 경우의 예상 가치는 적발 확률과 벌금에 의해 달라지기에 Allingham and Sandmo(1972)는 적발확률의 증가는 항상 더 큰 소득신고로 이어질 것이라고 말했다. 이론연구 이후

이러한 실증연구에서도 적발확률이 납세순응도를 높이는 것으로 나타났다. Pommerehne and Weck-Hannemann(1996; Weck-Hannemann and Pommerehne, 1989)은 스위스의 25개 주를 비교하였을 때 적발이 더 많이 발생한 주에서 납세 준수율이 높은 것을 확인하였으며, Ali et al.(2001)의 미국 국세청 데이터를 활용한 연구에서도 유사한 결과가 나타났다. Slemrod et al.(2001)은 실험 그룹에 우편으로 사전에 감사를 예고한 경우 저소득층과 중산층은 전년 대비 납세를 늘린 반면, 고소득자들은 납세가 줄어드는 것으로 나타나 적발확률과 소득의 반대 상호작용이 있음을 발견했다. 대체로 적발확률이 높아질수록 적발확률에 대한 정보가 정확할수록 납세순응도도 높아지는 것으로 나타났으나, Friedland(1982)는 반대로 적발확률이 낮을 경우에는 적발확률에 대한 정보가 부정확할 때 오히려 납세 준수율을 향상시킬 수 있다고 하였다. 이는 적발확률이 낮은 경우 실제로 적발당하는 경우가 많지 않기 때문에 납세자들이 적발확률에 대해서 실제보다 과도하게 평가하여 행동에 반영하기 때문이라고 할 수 있다.

반대로 Gërzhani and Schram(2006)은 알바니아와 네덜란드에서 실험을 수행하였는데, 알바니아 참가자들은 적발확률에 영향을 받지 않았으나 네덜란드 참가자들은 적발확률이 낮을 때 오히려 더 많이 납세를 회피하는 것으로 나타났다.

Fischer et al.(1992)은 인지된 확률과 객관적인 확률이 크게 다를 수 있음을 지적하며 주관적인 적발확률에 따른 세금 연구에 더 큰 관심을 보였다. 실제로 설문조사에 따르면 탈세를 인정한 납세자가 인식하는 적발확률이 정직한 납세자보다 더 낮았다(Mason and Calvin, 1978). 또한 주관적인 적발확률은 이전 경험에 따라 달라질 수 있다. Spicer and Hero(1985)는 10개의 기간에 걸쳐 무작위로 감사를 실시하였는데, 열 번째 기간의 납세 준수는 이전 감사 횟수와 상관관계가 있는 것으로 나타났다. 적발 경험이 적발확률을 평가하기 위한 학습과정을 유도한다는 것은 Guala and Mittone(2005)과 Mittone(2006)의 실험 연구에서도 알 수 있는데, 전체 60개 기간 동안 세금을 납부하고 첫 번째 기간에만 또는 마지막 30개 기간에만 감사를 받았을 때, 초기 감사를

받은 실험자의 납세순응도는 상대적으로 높았으며 전체 실험 기간 내내 높은 준수율을 보였다. 그러나 전반부에 감사를 받지 않은 참가자는 후반 30개 기간에 감사가 시작되어도 순응도가 높아지지 않았다.

4) 처벌 강도에 따른 (납세순응의) 변화

처벌(벌금) 강도는 적발확률과 더불어 탈세의 기대 이득을 결정짓는 매개 변수다. 처벌 강도와 적발확률은 곱의 관계로 서로 대체 가능하며, 이론적으로는 높은 벌금이 탈세를 더욱 위험하게 만들어 탈세를 억제한다. 하지만 실증연구에서는 벌금으로 인한 탈세 억제 효과를 항상 뒷받침하지 않는 것으로 나타났다. 관측된 벌금의 효과는 이론에서 기대하는 것보다 약하고, 심지어 어떤 연구에서는 벌금을 높이는 것은 탈세를 더 높이는 역효과를 가져오는 것으로 나타났다.

처벌이 탈세 억제에 효과가 있다고 밝혀진 실증연구로는 Friedland et al.(1978)과 Park and Hyun(2003)의 연구가 있다. 이들의 연구에서는 적발 확률이 납세순응에 미치는 영향보다 벌금이 더욱 강력하게 영향을 미치는 것으로 나타났다. 반면 Friedland(1982)는 적발확률이 벌금보다 납세순응에 더 영향을 미친다고 주장하였다. 이와 관련하여 Alm et al.(1995)는 벌금은 높은 적발확률과 조합해야만 효과가 있다고 주장하였다.

반면 처벌의 효과를 반증하는 연구 결과도 있다. Pommerehne and Weck-Hannemann(1996)은 스위스 주(canton)²⁹⁾별 조세순응을 비교하는 연구에서 처벌이 아무런 효과를 미치지 못함을 발견하였다. Ali et al.(2001)은 1980~1995년간 미국 납세자의 순응행태를 연구하였는데, 해당 기간 탈세에 대한 벌금 비율을 5%에서 30%까지 인상하였다. 연구 결과, 인상한 벌금 비율은 조세순응에 아무런 영향을 미치지 않았다.

실험적 연구를 진행한 Baldry(1987)와 Webley et al.(1991)의 연구에 의하면 실험이라는 도박과 같은 특수한 상황에서, 처벌의 영향이 실제보다 크게 나타나야 함에도 납세순응은 처벌의 영향을 받지 않은 것으로 나타났다. 또

29) canton: 스위스의 가장 작은 단위의 행정구역.

인지적 영향으로서 처벌의 영향을 연구한 Spicer and Lundstedt(1976)는 탈세 경향은 인지된 처벌의 심각성과는 관련이 없다는 결론을 내렸다.

흥미롭게도 처벌 강도를 높이는 것이 오히려 세금 회피를 늘리는 역효과를 나타내기도 한다. Schwartz and Orleans(1967)은 미국의 납세자를 대상으로 탈세에 대한 처벌 강화 내용을 담은 편지를 보내어 납세순응행태를 연구하는 실험을 진행하였다. 그 결과, 전년 대비 조정 후 총소득(adjusted gross income)은 증가하지 않았으나, 공제금(deduction)은 상당히 증가하였다. 이러한 결과는 납세자들이 처벌 강화로 인해 어쩔 수 없이 신고해야 했던 높은 소득에 대해 최대한 돌려받으려 노력했기 때문으로 해석할 수 있다.

5) 기대효용 이론이 설명하지 못하는 요인들

한편 이론이 납세순응 행위에 대해서 설명하지 못하는 부분들이 많자, 경제적 요소 이외에도 Tax morale이 납세순응에 미치는 영향을 분석하는 연구들이 많이 등장하게 된다. 특히 이는 납세순응에 여러 가지 요인들을 상대적으로 쉽게 분리할 수 있는 실험경제학 분야에서 발견되는 의문점과 연관성이 있다. 납세순응에 관한 실험들을 보면 실험참가자들은 이론의 예측보다는 좀 더 납세 순응적인 것으로 확인되었다(Alm, 1998). 이러한 현상을 파악하기 위해 실험참가자들이 낮은 탈세 적발확률을 과대 해석한다던가 혹은 극도로 위험 회피적인 성향이 있기 때문일 수도 있다는 가설에 대한 검증도 시도되었지만, 상당수의 실험 참가자들은 적발확률이 전혀 없는 경우에도 납세순응을 하는 모습을 보여 이러한 납세순응의 요인이 무엇인지에 대한 의문을 제기하게 된다(Alm et al., 1992).

이처럼 Allingham and Sandmo(1972) 모형만으로 설명되지 않는 납세순응 행위를 문화, 제도, 규범, 심리 등을 통해 설명하려는 시도들은 활발히 진행되고 있으며, 납세순응을 높이기 위한 정책 등에도 활용되고 있다(김재진, 2017; 박명호, 2016). 납세순응에 미치는 경제적인 요인 이외의 비경제적인 요인을 명확하게 구분하는 것은 어려운 일이다. 하지만 대표적으로 제

시되는 비경제적인 요인들은 개인적·사회적 규범, 조세형평성, 정부에 대한 신뢰 및 민주성으로 요약될 수 있다.³⁰⁾ 한편 Allingham and Sandmo(1972) 모형에서 다루어지지 않은 경제적인 요인인 납세협력비용 등 또한 납세순응에 미치는 요인 중의 하나로 해석되고 있다.

6) 규범

가) 개인적 규범

납세의 관점에서 개인적 규범(personal norm)은 ‘반드시 준수해야 할 도덕적 의무가 있다는 믿음’으로 정의될 수 있다(Wenzel, 2004; 홍성훈 외, 2017 재인용). Kirchler(2007)는 개인적 규범이 세금에 관하여 납세자의 가치와 윤리를 반영한다고 말했다. Torgler(2011) 등의 연구에서는 ‘납세윤리(tax morale)’를 ‘납세에 대한 내재적 동기(the intrinsic motivation to pay taxes)’, ‘납세를 기꺼이 하는 마음상태(taxpayers’ willingness to pay taxes)’, ‘납세순응에 대한 마음가짐(an attitude regarding tax compliance)’ 등으로 정의하고 있으며, 이러한 동기는 내적 및 외적 요인에 의해 영향을 받을 수 있다고 주장했다(박명호, 2015 재인용).

내적 요인과 관련된 대표적 연구로 Frey(1997)의 연구를 들 수 있는데, 그는 납세윤리(또는 납세의식, tax morale)를 본질적 종류의 내적 동기라고 보았다. 그는 무임승차 방지 및 공정성과 형평성을 달성하는 정책이 납세윤리(또는 납세의식, tax morale)를 지키는 수단이라고 하였다. 또한 Dwenger et al.(2016)은 당시 독일의 지방 교회세 납부가 순전히 자발적이었음을 이용하여 납세순응을 결정하는 요인으로서 본질적(내재적) 동기와 외적 동기(금전적 또는 벌칙 등의 외부 요인)를 연구하였다. 세금의 중요성과 세무조사 가능성 여부, 그리고 도덕적 가치에 대한 호소를 담은 편지를 각각의 그룹에 전달하였고, 그 결과 본질적(내재적) 동기에 반응하는 집단만 납세순응

30) OECD(2010)는 납세순응에 미치는 요인들을 억제(Deterrence), 규범(Norms), 기회(Opportunity), 형평성과 신뢰(Fairness and Trust), 그리고 경제적 요인(Economic Factors)라고 정리하였으며, 박명호(2016)는 개인적 규범, 사회적 규범, 조세형평성, 조세이해도, 국세청 신뢰, 탈세적발확률, 처벌 강도, 납세 협력비용을 납세순응의 결정요인으로 보았다.

에 긍정적인 영향을 미쳤다. Slemrod(1998)는 자발적인 세금 지불 의사로부터 파생된 사회적 자본은 행정 비용을 낮추고 시민들이 비용을 공평하게 부담하도록 한다고 주장했다. 이러한 관점들을 종합하여 홍성훈 외(2017)는 개인적 규범이 장기적으로 납세순응 전략의 핵심이 될 수 있다고 평가했다.

나) 사회적 규범

Knack and Keefer(1997)는 29개국을 대상으로 시민의 의무와 신뢰가 성장률과 투자율에 미치는 영향을 연구하였는데, 그 결과 사회적 자본 변수와 성장률 사이에 강력하고 유의한 양(+)의 관계가 도출되었다. 사회적 규범의 정확한 정의를 내리는 것은 어려운 일이나, OECD(2010)에서는 이를 사회 구성원이나 사회 계층 간의 행동이나 생각 또는 신념으로 정의하였다(박명호, 2015 재인용). 또 Elster(1989)는 사회적 규범이란 일정한 행동 양식이며, 이러한 행동 양식은 타인과 공유되고, 타인의 찬성과 반대를 통해 유지된다고 말했다. Fehr and Gächter(1998)는 사회적 규범이란 개인이 어떻게 행동해야 하는지 사회적으로 공유된 믿음에 기초하여 비공식적인 사회적 허가에 규정된 행동을 유발하는 행동 규칙이라고 정의하였다. 기존 문헌에 기초하여 박명호(2015)는 사회적 규범이 납세순응 행위에 영향을 줄 수 있다고 주장하였다.

Torgler(2002)는 재정 시스템은 비슷하지만, 사회적 규범이 서로 다른 여러 국가에서 조세순응 행태가 어떻게 나타나는지를 연구한 기존 문헌³¹⁾을 분석하였다. 주요 결과로 ① 조세에 순응하는 개인은 탈세를 부도덕한 행위로 보는 경향이 있음 ② 납세자에게 도덕적 호소를 한 경우, 순응도가 높아짐 ③ 지인 중에 탈세자가 있을 경우, 탈세할 가능성이 높아짐 ④ 사회적 결속이 강한 사회일수록 순응도가 높아짐 등이 있다.

정부가 사회적 규범에 영향을 줄 수 있다면 납세순응도가 높아질 것이라는 관점에 따라 Alm et al.(1999)은 정책을 통해 사회적 규범에 영향을 미치는 방법에 관해 연구했다. 이들은 개인이 조세순응을 결정할 때 영향을 받는

31) Alm et al.(1995); Yankelovich et al.(1984); Vogel(1974); Lewis(1979); De Juan et al.(1993)

사회 규범이 존재하고, 이러한 규범은 재정 시스템의 여러 측면에 대한 투표에 의해 영향을 받을 수 있다고 주장하였다. 또한 조세순응에 대한 사회적 규범은 집단 소통을 통해 영향을 받을 수 있다고도 하였다. Bohnet and Frey(1994)는 소통을 통해 집단의 결정을 개인화하며, Frey and Eichenberger (1999)는 더 나아가 사람들이 대화를 통해 자신의 선호를 충족할 방법을 찾고, 다른 사람들의 입장에 따라 자신의 입장을 깨닫게 된다고 주장하였다.

Bosco and Mittone(1997)은 집단 비난 감정(집단적 도덕적 구속)과 사회 후생 감소가 납세에 피해를 준다는 사실을 아는 것(주관적 도덕적 구속)이 탈세 행위를 감소시키는지 실험을 통해 검증하였다. 그 결과 두 가지 모두 탈세를 회피하게 만드는 강력한 도구로 나타났다. 그러나 Bosco and Mittone (1997)의 실험은 정태실험(1회성 실험)으로 한계가 명확했으며, Mittone(1997)은 이를 보완하기 위해 동태실험(반복실험)을 계획하여, 주관적 도덕적 구속이 탈세를 감소시킨다는 최종 결론을 얻었다.

Alm et al.(1995)은 실험을 통해 사회적 규범이 납세순응에 매우 중요한 결정 요소임을 밝혔다. 이들은 스페인과 미국에서 각각 진행한 실험의 결과를 비교하였는데, 정책 변화에 따른 순응도 변화의 정도가 서로 달랐다. 결국 이러한 차이는 납세순응에 대한 사회적 태도 - 납세윤리 또는 사회적 규범 - 에 기인한다고 결론 내렸다.

이와 더불어 도덕성, 사회적 규범 등을 강조하는 메시지를 통해 납세순응을 측정하고자 하는 방식은 현장 실험에서 자주 이용되었다. Torgler(2004)는 스위스 트리바흐(Trimbach) 지역을 대상으로 조세순응의 중요성과 자발적 납세를 독려하는 메시지를 보내는 실험을 진행했으나, 납세자들의 소득 신고와 세금 납부에 유의미한 영향을 발견하지 못했다. 한편 Fellner et al.(2013)은 호주 공중파 TV 및 라디오 수신료 납부 장려를 위해 영업장에 도덕 및 사회 규범적 메시지를 보내는 실험을 진행했다. 그 결과 모든 종류의 편지가 수신료 납부를 증가시키는 데 효과가 있었고, 규범적 메시지보다는 객관적 정보만을 담은 편지가 효과적으로 나타났다.

7) 조세형평성

조세형평성은 크게 수직, 수평, 그리고 교환의 형평성으로 나눌 수 있다. 박명호(2015)는 수직적 형평성이란 소득수준이 높은 사람들이 더 많은 세금을 부담하여 평균 세율이 소득에 대해 증가할 때, 수평적 형평성이란 소득수준이 유사한 사람들이 유사한 세부담을 질 때, 그리고 교환의 형평성이란 자신의 납세수준에 비해 정부로부터 공급받는 혜택이 클 때 각각의 형평성에 대해 긍정적으로 평가한다고 주장했다. 조세형평성에 대한 대부분의 연구는 수평적 형평성과 교환의 형평성에 초점이 맞춰져 있다.

먼저 형평이론(equity theory)에 기반한 수평적 형평성 연구를 진행한 Spicer and Becker(1980)는 실험 대상인 57명의 학생들에게 동일한 양의 돈을 주고 개별 세율은 40%라고 알려준 후, 19명에게는 평균세율이 65%, 또 다른 19명에게는 평균세율이 15%, 그리고 나머지 학생들에게는 모든 참가자의 세율이 같다고 말했다. 그 결과 총 예상세액 대비 평균 탈세 비율은 23.13%로 나타났으며, 집단별로는 자신의 세율이 평균보다 높다고 인지하는 그룹이 32%, 낮다고 인지하는 그룹이 12.26%, 그리고 세율이 같다고 인지한 그룹에서는 24.5%를 탈세했다. 이로써 이들의 연구 결과는 형평이론을 실증하며, 수평적 형평성이 납세순응에 영향을 줄 수 있음을 보였다.

반면 유사한 실험을 진행한 Webley et al.(1985)의 연구에서는 형평성이 탈세에 미치는 영향이 유의미하게 나타나지 않았다. 한편 Bott et al.(2014)이 노르웨이의 국외소득 납세자를 대상으로 진행한 우편 실험에서는 수평적 형평성을 강조한 편지가 납세자의 납세순응에 큰 영향을 미쳤다.

Torgler(2002)는 교환의 형평성이 높을수록 납세자의 법적 순응도가 증가한다고 하였다. Kim(2002)은 교환의 형평성과 납세순응에 대한 실험연구를 진행했는데, 정부로부터 공적이전(public transfer)을 받는 납세자는 이를 받지 않는 납세자에 비해 형평하다고 느끼는 것으로 나타났다. Becker et al.(1987)은 실험대상자들의 소득액과 관계없이 공적이전금(public transfer payment)을 지급하였는데, 이때 실험대상자에게는 지급 액수가 아닌 지급 비율을 알려준 후 세금 납부 실험을 진행했다. 그 결과 공적이전이 납세순

응에 중요한 역할을 하는 것으로 나타났다. 더 나아가 실험대상자는 자신이 받은 공적이전금(public transfer payment)이 다른 사람에 비해 적다고 느낄 때 탈세가 증가했다. 또 Pommerehne et al.(1994)은 동적 실험 방법을 사용하여 정부의 공공재 공급, 낭비, 공정성 고려와 납세자의 순응도 사이의 관계를 분석하였다. 실험 라운드가 진행될수록 사람들은 납세금액을 결정할 때 이전 라운드의 경험을 반영했다. 납세자의 공공재 최적선택과 실제 공급되는 공공재 수준의 차이가 벌어질수록, 과소 납세하는 동료 납세자가 많아질수록, 그리고 정부의 낭비가 심해질수록 개인의 세금 지불의사 금액이 줄어들었다.

이처럼 형평성과 관련된 대부분의 실험에서 교환의 형평성이 높을 때 납세자의 납세순응도가 높아짐을 확인할 수 있었다. 그러나 Alm et al.(1992)의 공공재 공급에 의한 소비자 잉여와 납세순응의 관계를 분석한 연구에서 소비자 잉여 증가와 납세순응도 사이에 비선형적 증가 관계가 나타났다. 이를 통해 교환의 형평성으로 정부가 납세자의 납세순응에 미칠 수 있는 영향에는 한계가 있는 것으로 드러났다.

8) 정부 신뢰 및 민주성

Torgler(2002)는 앞선 형평성 연구에서는 정부와 납세자 사이의 상호작용에 대한 영향을 간과하고 있다고 지적하며, 만약 납세자들이 직접 자신이 납부하는 세금의 활용에 관하여 투표할 수 있다면, 세금 납부에 더욱 적극적으로 될 것이라고 주장했다. 이와 더불어 Kidder and McEwen(1989)은 더 많은 사람들이 투표에 참여할수록 더 강한 의무감을 느낀다고 주장했다. 또 Daude et al.(2012)은 민주주의가 최고의 정치제도라고 생각하는 사람일수록 탈세를 정당화하지 않는다고 주장하였다(김재진, 2017 재인용).

민주성이 납세순응에 미치는 영향을 양적으로 연구한 Pommerehne and Weck-Hannemann(1996)은 주(canton)마다 시민의 정치참여 가능 수준이 다르고, 강력한 재정분권화가 이루어진 스위스를 실험의 장으로 삼아 시민의 정치 참여도와 납세순응의 관계에 관해 연구했다. 그 결과, 시민의 직접 정치

참여도가 높은 주일수록 납세순응도가 높았다. Feld and Kirchgässner(2000)는 이와 관련하여 직접 민주주의는 시민이 공동체에 대해 더욱 강력한 책임감을 느끼게 한다고 말했다. 또 Alm et al.(1993)은 정부가 투표 결과와 세금의 사용이 시민의 선호 방향과 동일하다는 사실을 시민들이 잘 인지하게 한다면, 더 큰 납세순응을 이끌어낼 수 있다고 주장하였다.

Torgler(2003)의 연구에서는 정부에 대한 신뢰와 시민으로서의 자부심이 납세 윤리(tax morale)에 유의미하고 긍정적인 영향을 미친다는 실증연구 결과를 제시하였다. 이와 더불어 납세 윤리(tax morale)는 국가와 의회에 대한 높은 신뢰와 종교적 믿음의 변수로 측정되는 시민의 도덕적 의무감으로 형성된다고 주장하였다(Torgler and Murphy, 2004).

한편 정부 신뢰에 대한 또 다른 측면의 연구로 과세관청과 납세자 사이의 공정성이 유발하는 신뢰에 관한 연구들도 있다. 이러한 연구들을 종합하면, 납세자의 자발적 순응을 향상시키기 위해서는 납세자에 대한 공정한 처우와 납세자-과세관청 사이의 상호 신뢰 및 협력이 중요하다(Braithwaite, 2011; Kirchler and Hölzl, 2006).

9) 납세협력비용

납세협력비용이란 납세자가 세법 및 과세관청이 요구하는 요건을 충족시키는 데 드는 비용이고, 실제 납부 세액을 초과하여 발생할 수 있으며 이에 따라 세금의 본질을 왜곡할 수 있는 비용으로 보고 있다(Sandford, 1995; 홍성훈 외, 2017 재인용). Sandmo(2005)는 심리적 요인을 중요시하는 행동적 이론을 수용하여 조세가 복잡할수록 납세자의 조세순응에 부정적 영향을 미치고, 처벌보다는 그로 인한 사회적 비용 증가를 고려해야 한다고 주장했다. Alm et al.(2010)은 실험을 통해 과세당국이 조세 관련 정보를 납세자에게 제공하면 납세순응도가 개선되는 결과를 얻었다. 이때 정보제공에 드는 비용보다 납세순응도 개선으로 증가하는 조세 수입이 더 크기 때문에, 기존의 처벌 방식보다 효율적이라고 할 수 있다. 또 Murphy(2004)는 납세자에 대한 공정하고 올바른 처우로서의 과세관청의 행정 처리(쉬운 법률용어, 손쉽게

접근 가능한 웹사이트, 손쉬운 세금납부, 자동과세, 개인정보 접근성, 세금제도의 간소화 등)가 과세관청에 대한 신뢰도 및 납세순응을 높일 수 있다고 주장하였다. 국내에서는 김기성 외(2014)의 실증연구에서 온라인 납세시스템의 편의성과 지방세제 공평성이 지방세 납세순응에 긍정적인 효과를 미쳤다는 결과를 얻었다. 납세협력비용과 관련하여 홍성훈 외(2017)는 조세회피를 어렵게 만들고, 동시에 납세순응을 쉽게 만들어 인지적 납세순응을 높이는 방안을 제시하였다.

10) 추가로 사회인구학 특성들이 미치는 영향

납세순응에 영향을 미치는 사회 인구학적 특성을 분류하자면, 납세자의 연령, 성별, 교육 수준 등을 포함하는 '납세자 개별 특성 요인'과 소득의 원천,³²⁾ 문화, 인구통계학적 특성 등의 요인으로 나누어 볼 수 있다. 우선 납세자 개별 특성 요인과 납세순응도 사이의 상관관계를 연구한 선행연구들을 살펴보면, 그 결과가 혼재되어 있음을 알 수 있다.

Spicer and Lundstedt(1976), Clotfelter(1983), Cummings et al.(2009) 등의 연구에서는 실험적 방법을 통해 납세자의 나이와 과거 경험 등이 납세순응에 영향을 미친다는 결과를 나타냈다. 반면 Dubin and Wilde(1988), Beron et al.(1992), Tauchen et al.(1993)의 연구에서는 나이 변수가 모든 납세자를 대변하는 변수가 아닐 수 있음을 지적하였다. 또 Porcano(1988), Collins et al.(1992) 등의 연구에서는 설명변수 간의 상관성 때문에 나이의 효과를 측정하기 어렵다고 주장했다.

홍성훈 외(2017)에 의하면 최근 경제적 주체로서 여성의 사회참여가 높아지면서 성별과 납세순응 사이의 직접 상관성은 사라졌다고 보고 있으며, 일반적으로 성별과 납세순응의 상관관계를 고려할 때 성별과 교육수준의 상관성을 분리하기 어려워 명확하게 제시된 결론이 없는 것으로 나타났다.

납세자의 교육수준에 대해서도 상반된 연구 결과가 존재한다. Torgler(2004)의 연구에서는 교육수준이 높을수록 납세에 긍정적인 결과를 얻었고,

32) 소득수준에 따른 납세순응도는 앞 부분을 참고.

Carroll(1992)은 세무 관련 교육수준과 납세순응 간 양(+)의 상관관계를 제시했다(홍성훈 외, 2017 재인용). 반면 국내에서는 교육수준이 낮을수록 순응도가 높다는 결론을 얻었다(박명호 · 심태섭, 2008).

납세자의 개별 특성 요인 외에도 납세순응에 영향을 미치는 요인을 분석한 시도는 여럿 있었다. Boylan and Sprinkle(2001)의 연구에서는 소득을 근로소득(earned income)과 증여소득(endowed income)으로 구분하여 각각의 원천에 따른 납세순응 행위를 분석하였다. 그 결과 근로소득의 경우 증여소득보다 세율 인상에 대해 더 많은 소득을 보고하는 것으로 나타났다. 이 밖에도 Richardson(2006)은 납세순응 연구에 문화적 요소가 고려되어야 한다고 주장했고, Tittle and Logan(1973)과 Trivedi et al.(2003)의 연구에서는 인구통계학적 특성이 납세순응에 통계적으로 유의미한 영향을 준다고 주장하였다.

Ⅲ. 텍스트를 이용한 연구 방법론 개관

본 장에서는 텍스트를 이용한 연구 방법론(텍스트 마이닝)에 대해 개괄적으로 소개하고 이후 분석에서 사용될 방법론에 관해서 설명하고자 한다.

1. 텍스트를 이용한 연구 방법론

텍스트를 전자적인 형태로 기록하는 기술이 발전함에 따라 언론의 보도와 정책입안자들의 의사록, 인터넷 미디어에서 사람들의 의사소통 등 방대한 양의 텍스트 자료에 접근하기가 쉬워지고 있다. 사회과학자들에게 이러한 텍스트 자료들은 기존의 연구들에서 사용되는 전통적인 데이터를 보완하는 유용한 데이터로 여겨지며, 경제학에서도 텍스트를 이용한 실증연구들이 활발히 이루어지고 있다(Gentzkow et al., 2019).

예를 들어 금융 분야에서는 뉴스나 소셜 미디어, 그리고 회사의 공시에 포함된 텍스트들을 이용하여 주가를 예측하거나 새로운 정보의 영향을 분석할 수 있다. 거시경제 분야에서는 검색엔진에서 사용자들이 입력한 텍스트를 이용해서 인플레이션이나 실업을 예측할 수 있으며, 뉴스 기사들을 분석하여 개별 국가들의 정부 정책의 불확실성을 측정할 수도 있다. 정치경제학 분야에서는 뉴스나 소셜 미디어들의 텍스트들을 정치적 편향을 연구하는 데 사용할 수 있으며, 정치인들의 의사록을 분석하여 정치적 의사결정 과정을 연구하는 데 사용할 수 있다. 산업조직론이나 마케팅 분야에서는 상품 리뷰에 대한 텍스트들을 통해 소비자의 의사결정을 이해하는 데 활용할 수 있다.

이처럼 텍스트 마이닝이라고 불리는 방법론은 여러 분야에서 광범위하게 사용되고 있기에 명확한 정의를 내리기 어렵고, 다양한 텍스트 마이닝 방법론을 모두 소개하는 것은 불가능에 가까울 것이다. 따라서 본 장에서는 텍

스트 마이닝 방법론을 소개하고 있는 Gentzkow et al.(2019)과 김수현 외 (2019)를 참고하여 텍스트를 이용한 연구 방법론을 개략적으로 소개하고자 한다. 이후 본 연구에서 사용할 연구 방법론에 대해서 구체적으로 서술하고자 한다.

텍스트 마이닝 방법론을 개념적으로 요약하자면 다음과 같은 과정으로 정리될 수 있다(Gentzkow et al., 2019).

1. 원본 텍스트인 문서 D 를 수량화된 배열인 C 로 나타낸다.
2. C 를 통해 알려지지 않은 관심변수인 V 의 추정치인 \hat{V} 를 추정한다.
3. \hat{V} 를 이용해서 기술적인 분석이나 인과관계 분석을 수행한다.

이상의 과정을 순서대로 자세히 설명하고자 한다. 우선 문서 D 는 분석의 대상이 되는 원본 텍스트 전체를 의미한다. 그런데 이 문서는 비정형화된 데이터이므로 이러한 문서를 수량화된 배열인 C 로 나타내는 과정에서 검토해야 할 사안들이 상당히 많이 존재한다.

우선 가장 첫 번째로 검토해야 할 사안들은 문서 D 를 개별 문서 $\{D_i\}$ 들로 어떻게 나눌까 하는 것이다. 이는 많은 경우에 예측을 하고자 하는 관심 변수인 V 의 특성들에 따라 결정된다. 만일 금융 분야의 뉴스를 통해 일별 주가를 예측하고 싶은 경우라면 개별 문서 $\{D_i\}$ 들은 일별로 나뉘는 것이 자연스럽다. 하지만 이러한 경우에도 같은 날짜의 같은 언론사의 뉴스들을 묶어서 하나의 개별 문서로 볼지, 각각의 뉴스들을 하나의 개별 문서로 볼지, 같은 섹터에 있는 뉴스들 묶어서 하나의 개별 문서로 볼지 등에 대해서는 명확한 이론적 기준이 존재하지 않는다. 따라서 이러한 판단은 연구자의 몫이며 여러 가지 분류 방법에 따른 연구 결과의 민감성에 대해서는 연구자가 충분히 검토할 필요가 있다.

다음은 개별 문서로 구분된 $\{D_i\}$ 들을 어떻게 수량화된 배열인 C 로 나타낼 것인가 하는 것이다. 개별 문서는 여전히 텍스트로 표현되어 있기 때문에 정량적인 분석을 위해서는 이를 수량화하는 작업이 필요하다. 우선 일반

적으로 이 과정에서 제일 먼저 하게 되는 작업은 단어가 아닌 요소들을 제거하는 과정이다. 이 과정에서 숫자나 문장부호들을 제거하게 된다. 또한 문장에서 매우 자주 사용되지만 큰 의미가 없는 단어들을 제거하는 과정을 거치게 된다. 이러한 단어들을 불용어(stop words)라고도 하는데, 대표적으로 영문에서는 'the', 'a' 등이 이러한 불용어에 포함되며, 한국어에서는 '것', '~할 수 있다'와 같은 표현들이 포함된다.³³⁾

문장부호와 숫자, 그리고 불용어를 제거한 후에는 문장들을 수량화하는 작업을 수행한다.³⁴⁾ 이때 텍스트를 수량화하는 방법에는 여러 가지 방법이 있을 수 있으나 대표적으로 단어 가방 모형(bag of words), n-gram, 단어 임베딩(word embedding) 등이 있다.

단어 가방 모형은 문장에서 나오는 단어들의 순서는 상관없이 단어들의 수만을 고려하는 단어 표상(word representation)이다. 예를 들어 개별 문서 D_1 이 'A B C A B A'라는 문장이라고 할 때, 해당 문장은 단어 A가 3개, 단어 B가 2개, 단어 C가 1개로 이루어져 있는 문장이므로, $C_1=(3, 2, 1)$ 로 표현이 가능하다. 단어 가방은 문서를 간편하게 수량화할 수 있다는 장점이 있지만, 단어의 순서에 대한 정보를 잃어버린다는 단점이 있다.

n-gram은 길이가 n인 구절을 수량의 단위로 표현하는 방법이다. 예를 들어 앞서 언급한 D_1 을 2-gram으로 표현하는 경우 AA가 0개, AB가 2개, AC가 0개, BA가 1개, BB가 0개, BC가 1개, CA가 1개, CB가 0개, CC가 0개이므로, $C_1=(0,2,0,1,0,1,1,0,0)$ 으로 표현 가능하다. n-gram 단어 간의 순서가 가지고 있는 정보를 일정 부분 유지할 수 있다는 장점이 있지만, 단어 가방에 비해서 단어의 표상이 더 복잡해지고 더 많은 연산 능력을 요구한다는 단점

33) Ranks NL company, 「Korea Stopwords」, <https://www.ranks.nl/stopwords/korean>, 검색 일자: 2022. 10. 4.; 길호현(2018).

34) 수량화하기 이전의 단계에서 TF-IDF(Term frequency-inverse document frequency)를 통해 매우 자주 나오는 단어나 매우 제한적으로 나오는 단어들을 추가로 제거할 수 있다. 추가로 어간추출(stemming)이라는 작업을 통해서 어간이 동일한 단어들을 묶어서 하나의 단어로 변환하는 작업을 거칠 수 있다. 이러한 작업은 수량화된 배열의 차원을 줄여주어 분석에 필요한 컴퓨터의 연산능력을 낮추어 주는 역할을 한다. 이에 대한 자세한 설명은 본 보고서의 범위를 벗어나므로 Gentzkow et al.(2019); 김수현 외(2019)를 참고하기 바란다.

이 있다.

마지막으로 단어 임베딩을 이해하기 위해서는 먼저 원-핫-인코딩(one hot encoding)에 대해서 이해해야 한다. 원-핫-인코딩은 단어 하나하나를 나타내는 이진적(binary) 벡터를 부여하는 것이다. 예를 들어 단어 A는 열벡터 {1,0,0}, 단어 B는 {0,1,0}, 단어 C는 {0,0,1}로 나타낼 수 있다. 이렇게 단어들에 벡터를 부여하면 문서들은 이 벡터들의 순서화된 배열로 나타낼 수 있게 된다. 앞의 예시에서 사용된 D_1 의 경우 $C_1 = \begin{Bmatrix} 100101 \\ 010010 \\ 001000 \end{Bmatrix}$ 라는 형태로 나타나게 되는 것이다. 이렇게 나타내는 경우 배열 C_1 의 크기는 전체 문서에서 나타나는 '어휘의 수(p) \times 문장의 길이(s)'가 되어 너무 차원이 큰 배열이 된다. 단어 임베딩은 이러한 C_1 의 차원을 p 보다 낮추어서 표현하는 방법들을 의미하며 대표적으로 Word2Vec 방법들이 있다. 단어 임베딩을 사용할 때는 단어들이 나타나는 위치의 정보를 모두 활용할 수 있으므로, 비슷한 위치에 나타나는 단어들은 서로 연관성이 있다는 것을 추론할 수 있게 하는 장점이 있다.

〈표 III-1〉 문서를 수량화된 배열로 나타내는 방법들

	$D_1 = A B C A B A$ $D_2 = B B C A C C A$
단어 가방 모형	$\{A B C\}$ $C_1 = \{3, 2, 1\}, C_2 = \{2, 2, 3\}$
2-gram	$\{AA AB AC BA BB BC CA CB CC\}$ $C_1 = \{0,2,0,1,0,1,1,0,0\}, C_2 = \{0,0,0,0,1,1,2,0,1\}$
단어 임베딩 (우측은 원-핫-임베딩이며 차원의 수를 필요에 따라 축소)	$A = \begin{Bmatrix} 1 \\ 0 \\ 0 \end{Bmatrix}, B = \begin{Bmatrix} 0 \\ 1 \\ 0 \end{Bmatrix}, C = \begin{Bmatrix} 0 \\ 0 \\ 1 \end{Bmatrix}$ $C_1 = \begin{Bmatrix} 100101 \\ 010010 \\ 001000 \end{Bmatrix}, C_2 = \begin{Bmatrix} 0001001 \\ 1100000 \\ 0010110 \end{Bmatrix}$

자료: 저자 작성

문서들의 수량화를 통해 배열 C 를 만들었으면 그다음 과정은 이 배열 C 를 통해 관측되지 않은 관심변수인 V 의 추정치인 \hat{V} 를 추정하는 과정을 수행

하게 된다. 이 과정에서도 여러 가지 방법론이 사용될 수 있으며 크게 사전 기반(dictionary-based) 방법론, 텍스트 회귀분석(text regression) 방법론, 생성 모형(generative model) 세 가지의 접근 방법으로 분류할 수 있다(Gentzkow et al., 2019).

사전기반 방법론은 통계적인 추론을 요구하지 않는 방법으로, 이미 구축된 사전을 통해 단어의 배열과 추정하고자 하는 값의 함수관계(f)를 설정하여, 문서 D 로부터 만들어진 수량적 배열 C 로부터 원하는 추정치 $\hat{v}_i = f(c_i)$ 를 얻어내는 과정을 의미한다. 예를 들어 금융 관련 뉴스에서 낙관적으로 사용되는 단어와 비관적으로 사용되는 단어들을 사전으로 구축한 다음, 낙관적인 단어의 수와 부정적인 단어 수를 c_i 로부터 계수하여 주식 가격에 대한 전망치 \hat{v}_i 를 만들어 낼 수 있다.

텍스트 회귀분석 방법론은 수량적 배열 C 를 독립변수로 생각하고 종속변수 V 를 추정하는 모형을 만드는 것이다. 이는 조건부 기댓값인 $E[v_i|c_i]$ 를 추정하는 과정으로 이해할 수 있으며 일반적인 회귀분석 방법론의 아이디어를 적용할 수 있다. 다만 이 경우에는 사전기반 방법론과는 달리 사전적인 함수관계 f 가 존재하지 않으므로, v_i 에 대한 관측치가 일부 있어야지만 $v_i = E[v_i|c_i] + \epsilon_i$ 의 함수관계를 추정할 수 있다.

또한 텍스트 회귀분석 방법론은 고차원의(high dimensional) 설명변수를 다루는 회귀분석 방법론을 사용하는 것이 필수적이다. 이는 수량적 배열 C 를 독립변수로 생각하는 경우 설명변수들의 차원은 해당 배열에서 사용된 단어의 종류에 의해서 결정되는데, 일반적인 텍스트에서 사용되는 단어의 종류의 수는 매우 많기 때문이다. 이러한 고차원 자료를 다루기 위해서는 모형 선택(model selection) 기능이 있는 벌점화 선형모형(Penalized linear model), 주 성분 회귀(principal components regression)와 같은 차원 축소모형, 서포트 벡터 머신(support vector machine)나 랜덤 포레스트(random forest)와 같은 비선형모형 등 여러 머신러닝(machine learning) 방법론이 사용될 수 있다.³⁵⁾

35) 머신러닝 방법론에 대한 이해와 구체적인 예시를 위해서는 이환웅 외(2021)를 참고하기 바란다.

생성모형은 텍스트 회귀분석과는 반대로 관심변수인 V 가 원래부터 존재하고, 이 V 로부터 텍스트 D 가 생성된 것이므로 수량적 배열 C 는 $p(c_i|v_i)$ 를 설명하는 확률적 모형을 통해 설명될 수 있다고 생각하는 방법론이다. 생성모형은 크게 비지도 생성모형(Unsupervised Generative Models)와 지도 생성모형(Supervised Generative Models)으로 구분될 수 있다.

비지도 생성모형의 상황에서는 연구자들은 관심변수의 실제 관측치 v_i 를 가지고 있지 않다. 이 경우에는 연구자들은 $p(c_i|v_i)$ 의 구조에 대한 강한 가정을 할 필요가 있다. 비지도 생성모형의 대표적인 사례는 토픽 모형화(topic modeling)가 있다. 토픽 모형은 관심변수 $v_i = \{v_{i1}, v_{i2}, \dots, v_{ik}\}$ 가 k 개의 토픽에 대한 가중치라고 가정하며, 이러한 가중치에 의해서 문서들이 형성된다고 가정한다. 이러한 토픽 모형을 사용하면 개별 문서들이 어떠한 토픽으로 구성되었는지 그 구성비를 나타내거나 혹은 개별 문서들이 속해 있는 토픽의 범주를 분류할 수 있다.

지도 생성모형은 v_i 에 대한 실제 관측치를 일부 가지고 있는 경우로 이를 이용해서 생성모형을 만들어 낼 수 있다. 지도 생성모형에서 가장 자주 활용되는 모형은 베이지 분류모형(Bayes classifier)이다. 이는 어떤 개별 문서가 v_i 라는 특성(예를 들어 스팸 메일인지 아닌지 여부)을 가지고 있을 때, 어떤 단어의 배열 C_i 가 나타나게 될 조건부 확률을 얻은 후 베이지 정리를 이용하여 역변환하여 어떤 단어가 나타나면 개별 문서의 특성이 어떻게 예측하는 것이다.

이상에서 논하였듯이 텍스트 분석 방법론에는 문서 D 를 C 로 수량화하는 과정, 그리고 수량적 배열 C 에서 관심변수인 v_i 를 추정하는 과정에서 매우 다양한 방법론들이 사용될 수 있다. 따라서 연구의 목적에 따라 적절한 텍스트 분석 방법론을 사용하는 것은 매우 중요하다. Gentzkow et al.(2019)은 텍스트 분석 방법론의 선택에 관해서 다음과 같은 가이드를 제시하였다. 우선 사전기반모형은 단어들과 추정하고자 하는 특성 간의 관계에 대해 사전(事前)적으로 명확한 정보를 가지고 있을 때 추정하고자 하는 참값인 v_i 가 어떤 경우에도 관측되기 어려울 때 사용할 수 있다. 예를 들어 정책의 불확

실성을 측정하고자 시도한 Baker et al.(2016)의 경우 정책의 불확실성을 측정할 수 있는 어떠한 신뢰할 만한 측정 방법이 없으므로, 사전기반 접근 방법을 통해 뉴스에서 uncertainty라는 단어가 나오는 경우를 계수하여 정책의 불확실성을 측정하였다. 텍스트 회귀분석의 경우에는 회귀분석의 장점을 가지고 있어서 실제로 관측된 v_i 수가 많을수록 관측치의 수(n)가 많아짐에 따라 추정치의 점근적 속성(asymptotic property)이 좋아지므로, 이런 경우에 사용하면 유용하다. 반면 지도적 생성모형의 경우에는 텍스트 회귀분석이 양호한 점근적 속성을 가지기 어려운 경우, 예를 들어 관측치의 수(n)가 차원의 크기(p)에 비해 압도적으로 크지 않은 경우에 유용할 수 있다. 마지막으로 비지도 생성모형은 연구하고자 하는 문서가 여러 가지 잠재적인 특성이 있다고 생각할 때 유용할 수 있다. 예를 들어서 어떠한 경제학 학술지의 문서를 연구하고자 한다면 미시, 거시, 계량, 금융, 노동, 산업조직 등 여러 가지의 주제를 다루고 있을 가능성이 있는데, 이러한 주제를 분류하기 위해서는 위에서 언급한 비지도 생성모형 중 하나인 토픽 모형을 사용하는 것이 유용하다.

이상에서 논의하였듯 텍스트 마이닝 방법론은 연산이 어려운 비구조적 텍스트 데이터를 수치화하여 구조화된 자료로 변환시킴으로써 정보의 원천으로 활용하게 하는 방법론을 모두 통칭하는 표현이라고 할 수 있다(김수현 외, 2019). 텍스트의 전자적인 저장능력과 컴퓨터의 연산능력이 좋아지고, 고차원 자료를 다루는 머신러닝 방법론이 비약적으로 좋아지는 만큼 텍스트 마이닝 방법론도 앞으로 여러 경제학 분야에서 많이 쓰일 것으로 예상된다.

2. 본 연구에서 사용할 텍스트 분석 방법론

서론(제I장)과 제II장 제2절에서 언급하였듯이, 본 연구의 핵심 연구 질문은 인터넷 미디어 여론이 조세정책에 대한 국민의 선호를 잘 반영하는가 하는 것이다. 만일 인터넷 미디어 여론이 국민들의 선호를 잘 반영한다면, 황의찬·우석진(2022)의 연구에서 밝힌 것처럼 조세정책에 대한 인터넷 뉴스 댓글의 감성지수가 낮아질수록 정부가 정책을 철회할 가능성이 커지는

것이 사회적으로 바람직할 수 있다. 하지만 만일 인터넷 미디어 여론이 국민들의 정책선호를 잘 반영하지 못한다면, 납세자들의 반발을 의식하여 이미 국회를 통과한 조세정책을 수정하는 것은 사회적 합의를 반복하는 것으로 공정하지 못하며 세제 안정성을 해친다고 할 수 있다.

인터넷 미디어 여론이 조세정책에 대한 국민의 선호를 잘 반영하는지 질문에 답하기 위해, 위에서 언급한 Gentzkow et al.(2019)의 이론적 틀을 차용해 보도록 한다. 이 이론적 틀에 의하면 사회적으로는 어떠한 참값들이 존재한다. 예를 들어 어떠한 조세정책에 대한 개인 i 의 선호가 v_i^T (ex. -1,0,1: 반대, 중립, 찬성)라고 존재한다면 $E(v_i^T)$ 는 사회적 선호라고 할 수 있다. 이 v_i^T 나 $E(v_i^T)$ 를 관찰할 수 있으면 좋겠지만 이는 투표와 같은 과정이 없다면 관찰되기 어려우므로, 정책결정자들은 텍스트로 나타나는 여론을 통해서 이에 대한 추정치를 얻고자 시도할 수 있다.

만일 조세정책에 대한 정확한 여론이 텍스트인 D_v 로 존재한다고 가정해 보자. 그렇다면 D_v 를 수집한 연구자는 이를 통해 C_v 라는 수학적인 배열로 표현할 수 있을 것이다. 그리고 C_v 를 통해서 \hat{v}_i^T 나 혹은 $E(\hat{v}_i^T)$ 을 추정하고자 시도할 수 있을 것이다.

하지만 현실에서 얻어지는 조세정책 관련 인터넷 뉴스 댓글과 같은 미디어 여론 (D_M)이 반드시 조세정책에 대한 정확한 여론인 D_v 와 같거나 유사하다고 볼 수는 없다. 인터넷 뉴스 댓글들은 정책에 대한 의견과는 별도로 현 정권에 대한 정치적 성향 등을 통해 발현될 수도 있고, 극단적으로는 댓글 아르바이트와 같이 경제적인 유인을 통해서 발현될 수도 있다. 따라서 본 연구에서는 정확한 여론인 D_v 와 미디어 여론 D_M 가 얼마나 유사한지 텍스트 분석 기법을 통해 분석해 보고자 한다. 그리고 이를 통해서 인터넷 댓글과 같은 미디어 여론이 조세정책 결정에 영향을 미치는 것이 과연 바람직한지, 아니면 오히려 왜곡되는 결과를 가지고 오는지 시사점을 도출해 보고자 한다.

연구 방법의 흐름

1. 우선 조세정책에 관한 정확한 여론을 포함하는 문서 $D_{v,r}$ 를 얻어야 한다. 본 연구에서는 설문조사에서 조세정책에 대한 선호(긍·부정)와 이에 관련된 의견을 개방형 질문을 통해서 수집하여 조세정책에 대한 여론을 포함하는 문서를 수집하였다. 물론 설문은 모집단이 아니라 샘플이기 때문에 한계점이 있기는 하지만, 실제로 응답자들의 선호 v_i^r 를 관찰할 수 있고, 이와 연결된 개별 문서인 $D_{v_i^r} \in D_{v,r}$ 를 얻을 수 있다.
 - 1.1. 설문조사를 수행하여 얻어진 응답이 조세정책에 관한 여론을 대표할 수 있는 신뢰할 만한 응답인지 확인하기 위해, 본 설문조사에서 얻은 응답이 선행연구들에서 얻어진 조세정책에 대한 선호들이 예측하는 응답과 유사한지를 검토하여 설문 응답의 타당성을 검토하였다(Sanity check)(제IV장 제1절).
2. 조세정책을 다루는 인터넷 뉴스 및 댓글 등을 통해서 조세정책에 관한 $D_{M,r}$ 를 수집한다. 본 연구에서는 네이버 뉴스 기사, 네이버 뉴스 댓글, 유튜브 댓글을 수집하였다.
3. 설문을 통해서 얻은 $D_{v,r}$ 와 인터넷 댓글 등을 통해 $D_{M,r}$ 를 직접적으로 비교하는 것은 어렵기 때문에 텍스트를 수량적 배열로 변환하여 $C_{v,r}$ 와 $C_{M,r}$ 를 만든 후 이 둘을 정성적으로 비교한다. 이러한 비교에 있어서는 연관규칙 학습과 의미망 분석 방법이 사용되었다(제IV장 제2절).
4. 3에 이어서 정량적인 방법으로 분석을 수행하였다. 우선 앞에서 설명한 비지도 생성모형 중 하나인 잠재 디리클레 할당(Latent Dirichlet Allocation, LDA)모형을 통해 토픽 모형화를 수행하여 $C_{M,r}$ 의 토픽의 구성을 파악하고, 그중에서 조세정책 관련 토픽이 차지하는 비중을 분석한다(제IV장 제1절).³⁶⁾
5. 마지막으로 $C_{v,r}$ 를 통해서 긍정과 부정의 감성을 분류하는 분류기를 만들어 낸다. 그리고 $C_{M,r}$ 를 분류기에 넣어서 감성의 위치를 분석한다. 이렇게 하면 다음 예시와 같은 결과물을 얻게 된다(제IV장 제2, 3절).

부정					긍정					
0	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
			네이버 댓글	유튜브 댓글			설문의 감성 ($C_{v,r}$)			

이상 일련의 과정을 통해서 인터넷 뉴스 댓글과 설문에서 얻어진 조세정책 관련 텍스트가 얼마나 유사한지 정성적·정량적으로 평가해 본다. 이를

36) LDA 모형은 비지도 생성모형이므로 해당 모형을 통해 $C_{M,r}$ 을 구성하고 있는 토픽의 구성을 분석할 때 조세정책 관련 토픽이 반드시 존재해야 하는 것은 아니다. 하지만 데이터의 수집과정에서 조세정책에 관련된 문서를 수집하였으므로, 조세정책 관련 토픽이 일정 부분 존재할 것으로 가정한다.

통해 인터넷 미디어 여론이 조세정책에 관한 국민들의 선호를 얼마나 잘 반영하는지, 그리고 인터넷 미디어 여론에 의해 정책 의사결정이 영향을 받는 것이 바람직한지 시사점을 도출해 본다.

IV. 설문과 인터넷 미디어 여론 텍스트의 정성적 비교

본 장에서는 설문을 통해 수집한 조세정책에 대한 여론과 인터넷 미디어에서 수집한 조세정책 관련 여론을 정성적으로 비교해 보고자 한다. 우선 조세정책에 대한 설문자료의 수집은 다음과 같이 진행되었다. 설문은 2022년 5월부터 10월까지 총 8차례 이루어졌으며 일반 국민들을 대상으로 온라인으로 수행되었다. 제1차부터 제5차까지의 설문에서는 가상자산 기타소득세, 금융투자소득세, 종합부동산세, 근로소득세, 탄소세에 대한 인식을 기존의 선행연구와 유사한 형태로 질문하였으며, 제6차부터 제8차까지는 가상자산 기타소득세에 대한 인식과 함께 공·부정에 관한 의견을 텍스트로 서술하여 응답하도록 하였다.³⁷⁾ 여러 조세정책에 대한 인식을 질문한 것은 기존의 선행연구에서 얻어진 예측들이 가상자산 기타소득세나 금융투자소득세 등 아직 도입되지 않은 조세에도 유사하게 인식되는지 확인함과 동시에 응답자들이 성실하게 응답을 하는지 파악하는 용도로 활용되었다(제Ⅵ장 제1절 참조). 가상자산 기타소득세에 대한 공·부정의 주관식 의견을 수집한 텍스트는 텍스트 분석 방법론을 통하여 인터넷 미디어 여론의 텍스트와 함께 정성적·정량적으로 분석하였다(제Ⅳ장 제2절, 제Ⅴ장 참조).

인터넷 미디어 여론의 텍스트는 가상자산 기타소득세에 대한 네이버 뉴스 및 댓글, 그리고 유튜브 댓글을 통해 수집하였다. 네이버 뉴스 기사 및 댓글 텍스트를 수집할 때 사용한 키워드는 ‘과세’를 반드시 포함해야 할 키워드로 하고 ‘암호화폐’, ‘가상화폐’, ‘가상자산’을 관련 키워드로, ‘규제’, ‘중부세’는 제외³⁸⁾할 키워드로 설정하였으며, 검색된 뉴스 기사와 해당 기사에 독자들이

37) 가상자산 기타소득세에 대한 동일한 설문을 여러 차수에 나누어서 진행한 것은 특정 시기에 발생하는 특정 이벤트의 영향을 최소화하고자 함이다.

38) 플랫폼은 ‘or’ 알고리즘을 바탕으로 검색하므로 여러 키워드로 검색할 경우 각 키워드에

작성한 댓글을 수집하였다. 가상자산소득 과세와 직접적 연관이 있는 기사는 2018년 이후부터 나타나므로 기간은 2018년 1월에서 2022년 6월로 하였다. 유튜브 댓글의 경우 2021년 가상자산소득 과세 방안 도입 및 연기와 관련한 7개 방송사(KBS, MBC, SBS, YTN, JTBC, TV조선, 채널A)의 유튜브 뉴스 게시물 중 댓글이 10개 이상 달린 게시물의 댓글 총 4,579개의 텍스트를 수집하였다.

1. 설문을 통해 수집된 조세정책 관련 여론 응답의 타당성 검토

본 절에서는 텍스트 분석 방법론을 통해 설문과 인터넷 미디어 여론의 텍스트를 비교하기 전, 설문을 통해 수집한 응답들이 신뢰할 만한 응답인지를 검토하고자 한다. 이는 설문의 응답들이 국민들의 실제 선호를 성실하게 잘 반영하는지를 간접적으로 확인하기 위한 작업으로, 설문의 응답들이 선행연구에서 얻어진 조세정책에 대한 선호들이 예측하는 응답과 유사한지를 검토한다. 텍스트 분석을 통한 설문과 인터넷 미디어 여론의 비교에 주된 관심 있는 독자들은 본 절을 생략하고 제Ⅳ장 제2절로 갈 것을 추천한다.

선행연구들은 납세자들의 납세의식³⁹⁾에 미치는 요인들로 경제적 부담, 민주적 절차, 교환의 형평성, 수직적 형평성 등이 있는 것을 보여주고 있으며, 사회·인구학적 특성들도 영향을 주는 것으로 알려져 있다(박명호, 2015). 본 설문의 제1차부터 제5차까지의 설문은 가상화폐 기타소득세, 금융투자소득세, 종합부동산세, 소득세, 탄소세에 대한 인식을 대상으로 수행되었으며, 해당 조세정책에 대한 긍·부정 의견과 경제적 부담, 민주적 절차, 교환의 형평성, 수직적 형평성과 같이 기존의 선행연구에서 납세의식에 미치는 것으로 알려진 요인들을 질문하였다. 또한 응답자들의 성별, 연령, 교육수준, 가구

관련된 기사가 모두 검색된다. 따라서 '과세' 키워드와 관련된 기사로 '중부세' 등에 관련된 기사가 있을 수 있으며, '암호화폐'에 관련된 '암호화폐 규제'에 관련된 기사들도 함께 검색된다. 이러한 기사들을 배제하기 위해 키워드 선정에 유의하였다.

39) 납세의식은 이미 도입된 조세정책에 대해서 적발이나 처벌이 없는 경우에도 얼마나 성실하게 해당 정책을 준수하고 세금을 납부할 의향이 있는지를 의미하는 것으로 조세정책에 대한 선호와 밀접한 연관이 있다고 할 수 있다.

소득, 이념적 성향, 직업, 보유주택 수, 거주지역 등 사회인구학적 특성을 수집하였다. 이렇게 수집된 응답은 선행연구의 방법론(박명호, 2015)과 유사한 분석을 수행하여 기존의 선행연구가 예측하는 결과를 가져오는지 확인한다. 한편 텍스트 분석을 위한 응답 자료는 제6차~제8차 설문을 통해 가상자산 기타소득세에 대한 긍·부정의 주관식 의견을 텍스트 형태로 수집하였다.

〈표 IV-1〉 선행연구에서 주로 사용되는 변수에 따른 긍정 비율

구분		가상자산 기타소득세	금융투자 소득세	종합 부동산세	근로소득세	탄소세
전체		61.2	50.8	61.0	54.1	62.2
경제적 부담	X	63.2	51.0	65.7	52.7	63.9
	-	54.8	50.2	60.4	55.0	65.4
	O	58.8	53.8	54.6	47.5	58.8
민주적 절차	X	34.4	31.1	44.8	30.8	43.4
	-	65.4	53.2	65.4	55.8	65.4
	O	84.5	78.6	76.1	72.8	83.2
교환의 형평성	X	48.0	41.2	52.2	40.8	51.7
	-	67.7	59.3	68.2	59.0	68.9
	O	81.9	73.2	74.5	70.8	83.8
개인적 규범	X	59.2	50.8	58.0	44.7	62.1
	-	55.0	51.4	58.8	46.5	58.0
	O	61.8	53.0	62.7	57.9	65.1
수직적 형평성	X	41.7	32.6	46.5	41.3	51.2
	-	56.3	52.4	56.1	50.6	59.3
	O	68.2	58.0	66.8	55.4	69.6

주: 긍정 응답을 100점, 부정 응답을 0점으로 환산

자료: 저자 작성

〈표 IV-1〉은 기존 선행연구에서 납세인식에 영향을 미치는 것으로 알려진 주요한 변수들에 따른 긍정 비율을 보여주고 있다. 경제적 부담(신규 세목의 경우 도입 시 예상되는 세금부담)이 낮다고 생각할수록 긍정 비율이 높게 나오지만 아주 큰 차이를 보이지는 않았다. 세금이 민주적 절차를 통해 도입·운용되고 있다고 생각할수록 긍정 비율이 높은 것을 확인할 수 있

있으며, 교환의 형평성(해당 세금을 내는 만큼 정부가 서비스를 제공함)이 높다고 생각할수록 긍정 비율이 높았다. 적발확률이 없어도 자발적으로 세금을 낼 것이라는 개인적 규범에 대한 응답이 높을수록 조세정책에 대한 긍정 비율이 높은 것으로 보이나 큰 영향을 미치지 않는 것으로 보인다. 한편 해당 조세정책으로 부자들이 더 많은 세금을 낸다고 하는 수직적 형평성에 대한 인식이 높을수록 조세정책에 대한 긍정 비율이 높았다. 이러한 결과들은 대체로 선행연구의 예측에 부합한다.

〈표 IV-2〉는 사회·인구통계학적 변수에 따른 긍정 비율을 나타내고 있다. 선행연구들의 분석에 의하면 사회·인구통계학적 변수들은 대체로는 납세인식에 큰 영향을 미치지 않는다는 것을 보이는데, 선행연구와 유사하게 사회·인구통계학적 변수들은 〈표 IV-1〉의 경제적·비경제적인 요인들에 비해 긍정 비율에 미치는 영향이 현저하게 적은 것으로 보인다. 하지만 선행연구에서 유의미하다고 나온 변수들과 본 설문에서 강한 영향을 주는 것으로 보이는 요인들의 비교를 통해서 본 설문의 내용이 이치에 맞는지(Sanity check) 확인해 보는 것은 의미가 있다. 선행연구에 의하면 자영업자들의 납세순응도가 높게 나오는데(박명호, 2016), 본 설문에서는 근로소득세에 대한 자영업자들의 긍정 비율이 높은 것으로 확인되었다. 다주택자의 경우 무주택자 대비 종합부동산세에 대해서 현저하게 낮은 긍정 비율을 보였으나, 다른 조세정책에 대해서는 큰 차이를 보이지 않는 것을 보인 것 또한 본 설문의 응답이 이치에 맞는 성실한 응답이었다는 것을 뒷받침한다. 한편 진보성향의 응답자가 보수성향의 응답자들보다 긍정 비율이 전반적으로 높으며, 금융투자소득세, 종합부동산세, 탄소세에서 그 차이가 큰 것으로 나타났다. 탄소세의 경우 10대는 높은 긍정 비율을 보이지만 20대는 낮은 긍정 비율을 보이는데, 이는 10대의 경우 장기적인 환경영향을 고려하지만 20대는 취업상황에 직면하여 거시적 실업률이나 기업의 국가경쟁력에 미치는 영향을 고려하기 때문으로 해석할 여지가 있다. 이상의 기술적인 통계를 보면 본 설문의 응답은 선행연구가 예측하는 결과 및 일반적으로 예상할 수 있는 결과에서 크게 다르지 않은 것으로 판단할 수 있다.

〈표 IV-2〉 사회·인구통계학적 변수에 따른 공정 비율

구분		가상자산 기타소득세	금융투자 소득세	종합 부동산세	근로소득세	탄소세
전체		61.2	50.8	61.0	54.1	62.2
성별	남자	62.3	50.6	61.3	56.9	61.9
	여자	60.0	51.0	60.6	50.7	62.5
연령	10대	57.0	49.3	58.9	50.0	70.7
	20대	53.7	45.6	53.2	45.9	55.8
	30대	60.2	50.5	62.7	53.6	63.0
	40대	59.7	51.3	61.8	51.9	61.7
	50대	67.2	53.0	66.4	55.9	63.9
	60대	67.6	54.0	59.9	67.3	61.0
교육 수준	중졸 이하	54.2	45.8	54.8	44.6	64.9
	고졸	64.5	56.7	64.8	57.1	64.8
	대졸	61.1	49.6	60.2	53.8	61.2
	대학원졸	57.5	47.2	59.9	52.8	62.3
가구 소득	300만원 미만	60.5	51.8	62.5	52.0	61.4
	600만원 미만	60.6	50.6	61.4	55.4	62.7
	600만원 이상	62.7	50.2	59.4	53.6	62.1
이념 적 성향	보수	55.6	44.5	53.5	53.9	56.6
	중도	61.4	50.3	60.5	52.2	60.6
	진보	65.3	57.7	69.0	60.8	72.8
직업	전문직	64.9	54.8	60.7	54.6	64.2
	자영업	61.8	54.0	63.1	62.5	60.6
	학생	56.5	48.3	57.9	52.7	68.4
	전업주부	68.0	57.1	62.3	55.7	65.8
	무직	58.5	49.7	56.5	55.7	52.6
	직장인	60.7	48.9	61.8	52.0	61.6
주택 보유 수	무주택	57.5	48.6	62.3	50.7	60.8
	1주택	64.3	52.9	62.6	56.3	62.5
	2주택	49.6	41.4	41.4	51.1	64.7
	3주택	65.3	50.0	50.0	54.2	69.4
지역	비수도권	59.7	49.8	60.3	52.2	62.7
	수도권	62.4	51.6	61.6	55.7	61.8

주: 긍정 응답을 100점, 부정 응답을 0점으로 환산

자료: 저자 작성

〈표 IV-3〉 긍정 응답에 대한 회귀분석 결과

VARIABLES	가상자산 기타소득세	금융투자 소득세	종합부동산세	근로소득세	탄소세
경제적 부담	-0.243*** (0.0610)	-0.167** (0.0685)	-0.338*** (0.0669)	-0.0641 (0.0763)	-0.180** (0.0755)
민주적 절차	1.161*** (0.0847)	0.926*** (0.0799)	0.710*** (0.0848)	0.747*** (0.0883)	0.849*** (0.0885)
교환의 형평성	0.171** (0.0828)	0.285*** (0.0789)	0.281*** (0.0812)	0.373*** (0.0836)	0.364*** (0.0854)
성실납부 의향	-0.00331 (0.0557)	0.0279 (0.0558)	0.0492 (0.0604)	0.172*** (0.0587)	0.107* (0.0586)
수직적 형평성	0.223*** (0.0676)	0.262*** (0.0715)	0.284*** (0.0719)	0.121* (0.0674)	0.306*** (0.0707)
Observations	800	762	702	701	715

주: 1. 종속변수는 긍정 응답을 1, 부정 응답을 -1로 설정함

2. Standard errors in parentheses

3. *** p<0.01, ** p<0.05, * p<0.1

자료: 저자 작성

〈표 IV-3〉은 긍정 응답에 대한 회귀분석 결과를 보여주고 있다. 경제적 부담에 대해서는 근로소득세를 제외하고는 모두 유의하게 음(-)의 결과를 보여주고 있다. 한편 민주적 절차와 교환의 형평성에 대해서는 모든 세목에 대해서 유의하게 양(+)의 결과를 보인다. 성실납부 의향은 유일하게 근로소득세에만 유의하게 양(+)의 효과를 보이는데, 이는 세목별 긍·부정 선호에 미치는 요인이 다를 수도 있음을 시사한다. 수직적 형평성은 모든 세목에 대해서 유의하게 양(+)의 결과를 보이나, 근로소득세의 경우는 10% 유의수준에서만 통계적 유의성이 있어 차이점을 나타내었다. 전반적으로는 선행연구에서 납세인식에 영향을 주는 것으로 알려진 요인들이 유의미하게 추정되었으며, 효과의 방향도 예상과 같았다.

〈표 IV-4〉 긍정 응답에 대한 회귀분석 결과(사회·인구통계학적 변수 포함)

VARIABLES	가상자산 기타소득세	금융투자 소득세	종합부동산세	근로소득세	탄소세
경제적 부담	-0.264*** (0.0630)	-0.162** (0.0701)	-0.290*** (0.0698)	-0.0342 (0.0791)	-0.193** (0.0766)
민주적 절차	1.181*** (0.0865)	0.950*** (0.0814)	0.715*** (0.0873)	0.731*** (0.0898)	0.849*** (0.0915)
교환의 형평성	0.209** (0.0864)	0.315*** (0.0826)	0.375*** (0.0837)	0.439*** (0.0865)	0.409*** (0.0886)
성실납부 의향	0.00270 (0.0573)	0.0292 (0.0578)	0.0316 (0.0624)	0.183*** (0.0614)	0.102* (0.0604)
수직적 형평성	0.219*** (0.0696)	0.241*** (0.0728)	0.282*** (0.0741)	0.113* (0.0686)	0.295*** (0.0716)
가구소득	-0.000216 (0.000380)	3.98e-05 (0.000385)	0.000523 (0.000396)	-0.000246 (0.000407)	5.64e-05 (0.000395)
주택 보유 수	-0.0516 (0.121)	0.112 (0.112)	-0.476*** (0.123)	-0.109 (0.127)	0.200 (0.123)
중졸 이하	-0.502 (0.421)	0.770 (0.473)	-0.925** (0.447)	0.123 (0.517)	0.622 (0.509)
고졸	-0.0357 (0.185)	0.291 (0.189)	-0.0767 (0.195)	-0.277 (0.198)	0.197 (0.196)
대학원졸	0.0861 (0.246)	0.222 (0.243)	0.423 (0.261)	0.0759 (0.242)	-0.0788 (0.244)
중도	-0.362* (0.198)	0.158 (0.196)	0.638*** (0.202)	0.114 (0.205)	0.0494 (0.201)
진보	-0.189 (0.240)	0.104 (0.239)	0.801*** (0.255)	0.0816 (0.249)	0.704*** (0.251)
20대	-0.179 (0.378)	0.584 (0.392)	-0.560 (0.423)	-0.172 (0.431)	-0.833** (0.401)
30대	-0.0944 (0.422)	0.647 (0.437)	-0.184 (0.483)	0.0624 (0.494)	-0.587 (0.440)
40대	0.111 (0.427)	0.886** (0.436)	0.238 (0.482)	-0.0144 (0.495)	-0.593 (0.443)
50대	0.365 (0.428)	0.864* (0.442)	0.376 (0.484)	0.340 (0.501)	-0.350 (0.445)
60대	0.466 (0.434)	0.782* (0.451)	0.347 (0.490)	1.056** (0.507)	-0.397 (0.456)
여성	-0.269* (0.148)	0.103 (0.149)	0.00441 (0.156)	-0.405** (0.158)	0.199 (0.158)

〈표 IV-4〉의 계속

VARIABLES	가상자산 기타소득세	금융투자 소득세	종합부동산세	근로소득세	탄소세
자영업	-0.286 (0.280)	0.144 (0.305)	-0.0250 (0.320)	0.253 (0.318)	0.299 (0.329)
학생	-0.555 (0.392)	-0.0537 (0.403)	-0.210 (0.437)	0.314 (0.398)	-0.506 (0.434)
전업주부	-0.288 (0.337)	0.186 (0.337)	-0.400 (0.364)	0.0481 (0.355)	0.410 (0.350)
무직	-0.146 (0.317)	-0.468 (0.347)	-0.309 (0.341)	-0.0661 (0.359)	-0.114 (0.345)
직장인	-0.340 (0.224)	0.0970 (0.230)	-0.0297 (0.246)	-0.112 (0.252)	0.285 (0.241)
Observations	800	762	702	701	715

주: 1. 종속변수는 긍정 응답을 1, 부정 응답을 -1로 설정함
 2. Standard errors in parentheses
 3. *** p<0.01, ** p<0.05, * p<0.1

자료: 저자 작성

〈표 IV-4〉는 사회·인구통계학적 변수를 포함하여 긍정 응답에 대한 회귀분석을 수행한 결과를 보여주고 있다. 사회·인구통계학적 변수를 통제하더라도 〈표 IV-3〉에서 제시된 결과와 크게 다르지 않아 사회·인구학적인 요인들은 대체로 유의미하지 않은 것을 확인할 수 있는데, 이는 선행연구에서도 동일하게 나타나는 결과이다(박명호, 2015; 2016). 한편 몇몇 사회·인구통계학적 변수는 통계적으로 유의하게 추정되었다. 중도 혹은 진보성향의 응답자는 보수성향 응답자에 비해 종합부동산세에 대한 긍정 응답 비율이 높았다. 진보성향의 응답자는 보수성향 응답자보다 탄소세에 대한 긍정 응답 비율이 높았다. 20대 응답자들은 10대 응답자보다 탄소세에 대한 긍정 응답 비율이 낮았다. 주택의 보유 수는 종합부동산세의 긍정 비율에 유의미하게 음(-)의 효과를 나타내었으나, 다른 조세정책에 대해서는 유의한 결과를 보이지 않았다.

이상의 회귀분석 결과들은 기존의 선행연구에서 알려진 납세인식에 영향을 미치는 요인들의 결과와 부합하며, 사회·인구통계학적 변수들이 긍정

응답에 미치는 영향도 이치에 맞는 것으로 볼 수 있다. 이는 본 설문이 전반적으로 성실한 응답으로 구성되었다는 것을 의미하며, 심한 편향성을 나타내고 있을 가능성을 현저하게 낮춘다고 할 수 있다. 따라서 본 설문에서 개방형 설문을 통해 질문한 가상자산소득 기타소득세 신설에 대한 선호에 대해서도 성실하고 진실하게 응답하였을 것으로 판단할 수 있을 것이다.

2. 텍스트 플랫폼별 조세정책 관련 여론 분석

본 절에서는 텍스트의 플랫폼별로 가상자산소득 과세 관련 여론을 텍스트 마이닝 방법론인 연관규칙학습과 의미망 분석을 통해 정성적으로 분석한다. 분석 대상 플랫폼은 네이버의 가상자산 기타소득 과세 관련 뉴스 기사와 그에 대한 댓글, 유튜브 뉴스에 대한 댓글, 그리고 앞에서 살펴본 설문조사 응답 내용을 포함한다. 본 절은 먼저 연관규칙학습과 의미망 분석 방법론을 먼저 소개한 후, 이를 이용하여 플랫폼별로 텍스트를 분석하고 이를 비교하는 순서로 구성된다.

가. 텍스트 마이닝 방법론

1) 연관규칙학습

연관규칙학습(association rule learning)은 Agrawal et al.(1993)이 제안한 방법론으로, 본래 슈퍼마켓 장바구니 데이터를 분석하여 상품 간 구매의 연관성을 뜻하는 ‘연관규칙’을 찾는 데에 이용되었다. 슈퍼마켓 데이터에서 연관규칙이란 특정 상품 집합을 동시에 구매하는 경향성을 뜻한다. 예를 들어 ‘된장, 호박, 두부’ → ‘감자, 양파’라는 연관규칙은 된장, 호박, 두부를 구매하는 경우 감자와 양파 또한 구매할 경향성이 높아짐을 의미한다.

장바구니 데이터를 이용한 연관규칙학습은 다양한 분야에 응용되고 있는데, 텍스트 마이닝의 경우 특정 단어의 집합들이 동시에 등장하는 경향성을 분석하는 데에 이용된다. 이때 장바구니 분석에서의 상품에 해당하는 단위는 단어이며, 장바구니에 해당하는 단위는 문장, 문단, 혹은 전체 글 등이 될

수 있다. 예를 들어 문장을 장바구니 단위로 하는 연관규칙학습 결과 ‘보호 → 환경’이라는 규칙이 도출되었다면, 이는 어떤 문장에 ‘보호’라는 단어가 포함되어 있을 때 ‘환경’ 또한 포함되는 경향성이 발견되었다는 뜻이다.

연관규칙학습 방법론이 장바구니 분석에서는 슈퍼마켓 이용객들의 효율적인 동선 설정 및 상품 배치 등에 이용된다면, 텍스트 분석에서는 단어의 나열인 텍스트로부터 어휘들 사이의 관계를 파악함으로써 최종적으로 특정 의미를 추출(semantic data mining)하는 데에 적용된다.

연관규칙학습에 이용되는 주요 개념은 통계의 확률 개념으로 설명할 수 있다. 우선 특정 단어 집합 A가 임의의 문장에 포함될 확률, 즉 P(A)를 ‘A의 지지도(support)’라고 한다. 예를 들어 A={환경, 보호}라면 A의 지지도는 한 문장이 환경과 보호 두 단어를 모두 포함할 확률이다. 다음으로 ‘B → A의 신뢰도(Confidence)’는 단어 집합 B가 속한 문장에 A 또한 포함될 확률로, 조건부 확률 P(A | B)를 통해 나타낼 수 있다. 마지막으로 ‘B → A의 향상도(Lift)’는 특정 문장에 B가 속해 있다는 정보가 A의 출현 확률을 향상시키는 정도를 뜻하며, 조건부확률과 비조건부확률의 비율 P(A | B)/P(A)로 나타낼 수 있다.

연관규칙학습을 통한 분석은 우선 지지도가 높은 고빈도 단어 집합을 확인함으로써 각 텍스트에서 자주 사용되는 어휘를 파악하고, 이후 신뢰도와 향상도가 높은 연관규칙을 통해 특정 어휘 집합들의 쓰임을 파악하고 그로부터 의미를 도출하는 과정을 거친다. 고빈도 단어 집합을 찾아내는 과정은 Agrawal and Srikant(1994)의 아프리오리(apriori) 알고리즘에 따르는데, 계산 과정에서의 효율성을 높이기 위함이다. 구체적으로 특정 텍스트에 사용된 전체 단어의 개수가 N개라면 가능한 단어 집합의 개수는 2^N 개이므로, 데이터가 커질수록 모든 가능한 단어 집합의 지지도를 계산하는 시간이 기하급수적으로 늘어나기 때문에 계산 시간을 줄이는 알고리즘이 필요하며 아프리오리 알고리즘이 이에 해당한다. 아프리오리 알고리즘은 고빈도 단어 집합의 모든 부분집합 역시 고빈도라는 점에 착안하여 원소가 k개인 집합에서 빈도가 낮은 집합을 제외하고 남은 집합의 초집합 중 원소가 k+1개인 집합을

고려하는 방식으로 계산시간을 줄인다.

아프리오리 알고리즘을 통하여 고빈도 단어 집합을 생성한 이후 해당 고빈도 집합 내에서 연관규칙을 생성한다. 구체적으로 각 고빈도 단어 집합을 2분할하는 부분집합 사이의 신뢰도가 일정 수준 이상인 경우, 두 부분집합이 연관규칙을 형성한다고 규정한다. 예를 들어 {보호, 환경, 기휘}라는 단어 집합이 고빈도 집합이라면 {보호, 환경} → {기휘}, {환경} → {보호, 기휘} 등 가능한 모든 분할의 신뢰도를 계산하여 신뢰도가 높은 규칙을 찾아낸다. 해당 과정을 거쳐 찾아낸 특정 신뢰도 이상을 보이는 부분집합의 관계를 연관규칙이라고 하며, 각 연관규칙의 강도는 향상도를 통해 측정한다.

한편 아프리오리 알고리즘에서 빈도가 낮은 집합을 제외할 때 이용되는 지지도의 문턱값과 연관규칙으로 인정되는 신뢰도의 문턱값을 정하는 과정에서 연구자의 재량이 개입한다. 문턱값을 높게 잡으면 계산이 빨라지고 자주 언급되는 어휘에 집중할 수 있게 되는 반면, 자주 언급되지 않지만 중요한 의미를 담고 있는 단어 집합 혹은 연관규칙을 분석에서 배제할 위험이 발생한다. 결론적으로 문턱값의 설정은 텍스트의 특성을 반영할 수밖에 없는데, 각 문장에서 사용되는 어휘가 제한적인 텍스트의 경우 지지도가 높게 나오는 단어 집합이 많기 때문에 지지도 문턱값을 높게 설정해야 계산과 이후 분석 과정이 쉬워지며, 반대로 어휘의 분포가 산발적인 텍스트의 경우 문턱값이 낮아야 분석 대상 단어 집합과 연관규칙이 풍부해지기 때문이다.

2) 의미망 분석

연관규칙 분석 다음으로는 의미망 분석을 진행한다. 의미망 분석은 연관규칙 분석 과정에서 도출된 고빈도 단어 집합을 의미망(semantic network)으로 시각화하여 단어의 클러스터 형성 여부 등 네트워크의 패턴을 분석하여 텍스트로부터 의미를 도출하는 방법론이다. 구체적으로 고빈도 단어 집합 중 크기가 2인 집합, 즉 어휘 두 개로 이루어진 단어 집합만을 이용하여 해당 집합에 속한 각 단어로 노드(node)를 표현하고, 같은 고빈도 단어 집합에 속해 있는 단어들 사이에 엣지(edge)를 연결하는 방식으로 의미망을 생성한다.

이렇게 생성한 의미망은 직관적으로 클러스터 형성 등을 분석하기에 쉬운 Fruchterman-Reingold force-directed 알고리즘으로 시각화하여 분석한다. 이때 각 노드의 크기는 해당 단어의 등장 빈도를 나타내며, 노드의 크기와 각 노드가 연결된 패턴 등을 분석하여 의미를 도출할 수 있다.

더불어 네트워크 분석에서 자주 사용되는 중심성(centrality) 분석 또한 의미망 분석에 사용될 수 있다. 중심성은 특정 노드가 네트워크에서 차지하는 위상을 나타내는 개념으로, 각 노드에 연결된 엣지의 수를 반영하는 연결 중심성(degree centrality), 최단경로 계산 시 각 노드가 차지하는 중요성을 반영하는 매개 중심성(betweenness centrality), 각 노드까지의 거리를 반영하는 근접 중심성(closeness centrality), 인접한 노드의 중심성까지 반영하여 계산하는 고유벡터 중심성(eigenvector centrality) 등의 중심성 개념이 존재한다. 이러한 중심성 분석을 통해 의미망에서 각 노드의 위상을 분석함으로써 텍스트에서 중요한 역할을 하는 어휘를 새로이 포착할 가능성 또한 시험해 보고자 한다.

한편 의미망 분석에 사용되는 고빈도 단어 집합이 앞서 연관규칙 분석 과정에서 생성한 고빈도 단어 집합과 동일하기 때문에 두 분석 결과의 차이가 크지 않을 수 있다. 하지만 의미망 분석을 통해 수많은 노드의 크기와 노드들의 클러스터 형성을 시각적으로 분석함으로써 단어 집합의 지도도와 연관규칙 분석만으로 포착하기 어려운 패턴을 효율적으로 파악할 수 있다는 점에서 해당 방법론을 추가로 도입한다.

나. 유튜브 댓글 분석

본 소절에서는 인터넷 댓글의 텍스트 분석 대상으로서의 적절성을 가늠해 보기 위하여 유튜브 뉴스 기사에 달린 댓글을 상술한 텍스트 마이닝 방법론을 통하여 분석해 본다. 2021년 가상화폐 소득 과세 방안 도입 및 연기와 관련한 7개 방송사(KBS, MBC, SBS, YTN, JTBC, TV조선, 채널A)의 유튜브 뉴스 게시물 중 댓글이 10개 이상 달린 게시물의 댓글 총 4,579개의 데이터를 이용한다.

인터넷 댓글을 분석할 수 있는 텍스트 데이터로 변환하기 위하여 웹스크레이핑 및 데이터 전처리의 과정을 거친다. 웹스크레이핑은 파이썬의 셀레니움(selenium)과 뷰티풀썬(beautiful soup) 모듈을 중점적으로 사용하였으며, 데이터의 전처리는 파이썬의 한국어 자연어처리 패키지인 코엔엘파이(KoNLPy)를 이용하였다. 댓글에 사용된 어휘 중 명사만을 선택적으로 분석하기 위하여 Okt(Open Korea Text) 패키지를 이용하여 품사 태깅을 진행하였다.

우선 <표 IV-5>에 아프리오리 알고리즘을 활용하여 생성한 지지도 0.02 이상의 고빈도 단어 집합을 지지도 순으로 나열한 결과를 제시하였다. 가장 빈도가 높은 단어 집합은 {세금}, {코인}, {가상화폐}, {보호}, {세금, 보호}로 각각 70%, 25%, 23%, 18%, 15%의 문장에서 발견되었다. 가상화폐 과세와 관련하여 빈번하게 사용될 것으로 예상되는 어휘들이 포착된 가운데 ‘보호’, ‘정부’ 등의 어휘 또한 발견되어 가상화폐 투자자에 대한 보호나 정부에 관한 의견들이 다수 있음을 파악할 수 있다. 단어의 개수가 두 개인 단어 집합 중에는 {세금, 보호}, {세금, 가상화폐}, {코인, 세금} 등 지지도가 가장 높은 ‘세금’이 포함된 집합이 역시 높은 지지도를 보인다.

<표 IV-5> 가상자산 과세 관련 유튜브 뉴스 댓글의 고빈도 단어 집합

단어 집합	지지도	단어 집합	지지도
{세금}	0.704	{정부}	0.084
{코인}	0.250	{세금, 가치}	0.081
{가상화폐}	0.225	{주식}	0.078
{보호}	0.180	{세금, 자산}	0.073
{세금, 보호}	0.150	{인정, 세금}	0.071
{과세}	0.147	{비트}	0.069
{왜}	0.136	{사람}	0.066
{돈}	0.136	{코인, 비트}	0.066
{세금, 가상화폐}	0.129	{정부, 세금}	0.060
{코인, 세금}	0.116	{말, 세금}	0.059
{가치}	0.114	{투자}	0.057
{자산}	0.110	{자산, 가상화폐}	0.056

〈표 IV-5〉의 계속

단어 집합	지지도	단어 집합	지지도
{돈, 세금}	0.107	{나라}	0.053
{왜, 세금}	0.107	{그냥}	0.053
{인정}	0.093	{국민}	0.051
{말}	0.091	{주식, 세금}	0.050

자료: 분석 결과를 바탕으로 저자 작성

다음으로 〈표 IV-6〉에는 신뢰도 0.7 이상의 연관규칙을 향상도 순으로 나열하여 제시하였다. 향상도가 가장 높은 연관규칙은 {무형} → {자산}으로, {자산}의 지지도(후향지지도)는 0.110으로 10% 수준이지만, {무형} → {자산}의 신뢰도(즉 조건부확률)는 0.786으로 높기에, 향상도는 $0.786/0.11=7.153$ 으로 나타났다. ‘자산’은 높은 빈도(지지도)로 나오는 단어가 아니지만, 문장에 ‘무형’이 포함되면 ‘자산’ 또한 포함될 확률(신뢰도)이 79%로 높아지기 때문에 {무형} → {자산}의 향상도가 매우 높은 것이다. 하지만 이는 ‘무형 자산’이라는 합성 명사를 Okt 태깅 패키지가 각각의 명사로 취급했기 때문이며, {무형} → {자산}이 중요한 의미를 지니는 연관규칙이기 때문은 아니다. 이는 분석의 컨텍스트에 맞추어 태깅 패키지의 사전을 업데이트할 필요성을 보여주며, 동시에 신뢰도나 향상도 해석에 있어 연구자의 재량이 들어가는 텍스트 마이닝 방법론의 정성적 특성을 보여준다. 특히 연관규칙학습은 정량적 분석과 정성적 분석이 모두 중요하게 적용되는 방법론임을 시사한다.

한편 본 학습 결과에서 더욱 의미 있게 해석할 수 있는 연관규칙은 {투자자, 세금} → {보호}와 {나라, 보호} → {세금} 등이다. 해당 키워드를 포함한 댓글을 보면 ‘나라’에서 ‘투자자’를 ‘보호’하는 데에 집중하지 않고 ‘세금’만 걷으려고 한다는 비판적인 시각이나, 가상화폐 소득 양성화를 통해 ‘세금’도 걷고 ‘투자자’도 ‘보호’해야 한다는 시각 등을 확인할 수 있다. 세법개정안에서의 가상자산소득 과세 관련 논의에서 투자자 보호 측면이 강조되지는 않는다는 점을 생각하면, 이는 댓글 분석을 통해서 발견할 수 있는 새로운 시각이다.

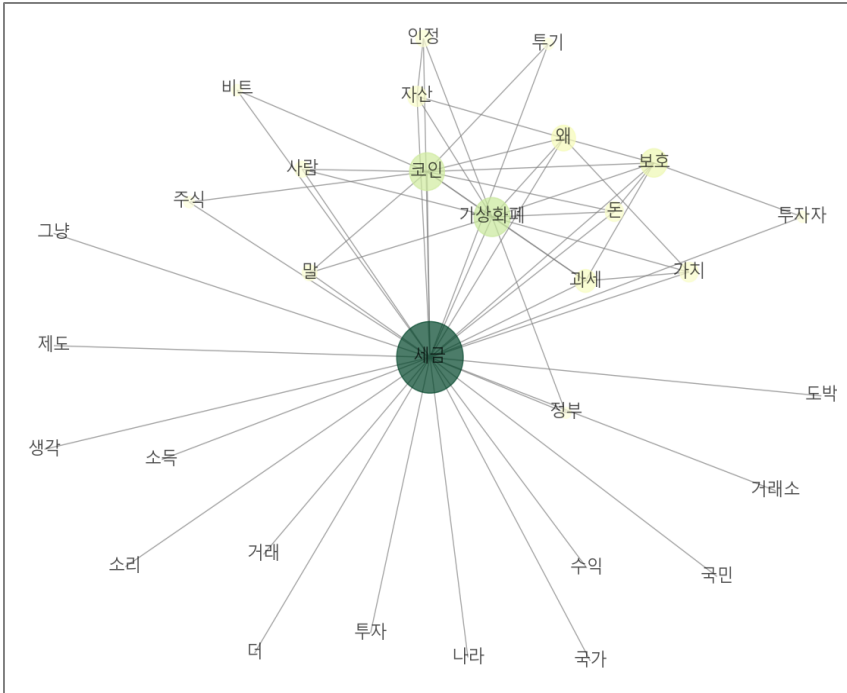
〈표 IV-6〉 가상자산 과세 관련 유튜브 뉴스 댓글의 연관규칙

전항	후항	전항지지도	후항지지도	신뢰도	향상도
{무형}	{자산}	0,017	0,110	0,786	7,153
{투자자, 세금}	{보호}	0,028	0,180	0,766	4,262
{비트}	{코인}	0,069	0,250	0,957	3,824
{비트, 세금}	{코인}	0,032	0,250	0,943	3,771
{비트, 가상화폐}	{코인}	0,013	0,250	0,857	3,427
{나라, 보호}	{세금}	0,014	0,704	1,000	1,419
{안내}	{세금}	0,013	0,704	1,000	1,419
{돈, 나라}	{세금}	0,011	0,704	1,000	1,419
{왜, 보호}	{세금}	0,030	0,704	0,961	1,364
{말, 왜}	{세금}	0,013	0,704	0,952	1,352
{부과}	{세금}	0,020	0,704	0,941	1,336
{도둑}	{세금}	0,016	0,704	0,923	1,310
{나라}	{세금}	0,053	0,704	0,921	1,308

자료: 분석 결과를 바탕으로 저자 작성

다음으로 고빈도 단어 집합을 이용한 의미망 분석 결과를 제시한다. [그림 IV-1]은 앞서 아프리오리 알고리즘을 통해 생성한 지지도 0.02 이상의 단어 집합 중 크기가 2인 집합을 모아 의미망으로 시각화한 그림이다. 앞서 지지도 분석에서 살펴본 바와 같이 ‘세금’, ‘가상화폐’, ‘코인’ 등의 노드의 크기가 큰 것을 확인할 수 있다. 한편 ‘세금’-‘가상화폐’-‘코인’-‘자산’-‘인정’ 등의 단어가 클러스터를 형성하고 있는데, 가상화폐 혹은 코인을 자산으로 인정하는 것이 가상화폐 과세 관련 논의와 밀접한 연관성을 가지고 있다는 인식을 확인할 수 있다. 또한 ‘투자자’-‘보호’-‘돈’-‘가상화폐’-‘과세’의 클러스터에서는 앞서 연관규칙 분석에서 본 바와 유사하게, 가상화폐 투자자들을 보호하는데 주력하지 않으면서 세금을 걷는 행태를 비판하는 의견을 확인할 수 있다. 가상화폐가 합법적 금융자산으로 인정되지 않음에도 가상화폐를 통한 금융 소득에 과세하려는 정책적 모순에 반응하는 의견이다. 가상화폐 과세를 둘러싼 논란에서 가상자산의 합법화 및 투자자 보호 방안 마련이 댓글을 작성한 사람들에게 중요한 쟁점임을 알 수 있다.

[그림 IV-1] 가상자산 과세 관련 유튜브 뉴스 댓글의 의미망 시각화



자료: 분석 결과를 바탕으로 저자 작성

마지막으로 <표 IV-7>에서는 다양한 중심성 개념을 이용하여 의미망에서 각 단어가 차지하는 위상을 분석하였다. 하지만 분석 대상 의미망의 구조가 복잡하지 않아, 서로 다른 중심성 개념에 따른 노드별로 중심성 순위의 차이가 크지 않은 것으로 나타났다. 주요 개념어 외에 '보호'의 중심성이 높게 나타난 것은 앞서 지지도와 연관규칙을 통해 분석한 결과와 유사한 것으로 나타나, 이 경우 중심성 분석에는 큰 추가적인 분석적 의의가 없음을 확인하였다.

〈표 IV-7〉 가상자산 과세 관련 유튜브 뉴스 댓글의 중심성 분석

단어	연결 중심성	고유벡터 중심성	매개 중심성	근접 중심성
세금	1	0.528	0.846	1
가상화폐	0.4	0.361	0.04	0.625
코인	0.367	0.329	0.036	0.612
보호	0.233	0.264	0.007	0.566
왜	0.2	0.243	0.003	0.556
과세	0.167	0.221	0.001	0.545
가치	0.133	0.18	0	0.536
돈	0.133	0.197	0	0.536
자산	0.133	0.169	0.001	0.536
인정	0.1	0.141	0	0.526
말	0.1	0.162	0	0.526
사람	0.1	0.162	0	0.526
비트	0.067	0.114	0	0.517
정부	0.067	0.118	0	0.517
주식	0.067	0.114	0	0.517
투자자	0.067	0.105	0	0.517
투기	0.067	0.114	0	0.517

자료: 분석 결과를 바탕으로 저자 작성

본 소절에서는 텍스트 마이닝을 이용하여 유튜브 뉴스 댓글 분석을 진행하였다. 댓글 분석은 분석 대상 의견을 빠르고 동시성 있게 파악할 수 있다는 큰 장점이 있지만, 댓글 텍스트에 잡음이 많아 분석 결과를 해석하는 데 연구자의 재량이 많이 동원된다는 점이 한계로 보인다. 따라서 빠르게 쟁점을 파악하여 추후 정교한 분석을 수행하기 위한 보조적이고 선제적인 의견 조사 수단으로서 댓글에 대한 텍스트 마이닝을 사용하는 것이 적절할 것으로 여겨진다. 다만 투자자 보호 관련 인식을 포착한 점에서 확인할 수 있듯이, 잡음을 효과적으로 제거한다면 의미 있는 패턴을 추출할 수 있는 여지 또한 존재한다.

다. 가상자산 과세 관련 네이버 기사 및 댓글 비교 분석

다음으로 기사와 댓글 텍스트의 차이를 알아보기 위하여 가상자산소득 과세 관련 네이버 기사 및 댓글 분석을 진행한다. 가상자산소득 과세와 직접적 연관이 있는 기사는 2018년 이후부터 나타나므로 기간은 2018년 1월에서 2022년 6월까지로 하였다. 최종 분석에 사용된 텍스트 자료는 뉴스 기사 832건과 이에 대한 댓글 3만 9,704건이 분석 대상에 포함되었으며, 뉴스 기사 중 가상화폐 과세 관련성이 높은 문장을 추출하기 위하여 ‘가상’, ‘화폐’, ‘암호’, ‘세금’, ‘과세’, ‘코인’, ‘자산’, ‘소득’ 중 한 단어 이상 포함된 문장만 분석하였다. 댓글은 모든 문장을 분석하였다. <표 IV-8>에 최종 분석 대상이 되는 기사 및 댓글의 문장 수를 제시하였다. 기사 1만 1,390문장, 댓글 5만 5,483문장이 분석에 포함되었다.

<표 IV-8> 연도별 분석 대상 문장 수

연도	기사 수	기사 문장 수	댓글 수	댓글 문장 수	기사별 댓글	언론사 수
2018	32	401	1,995	3,112	62.3	14
2019	55	900	2,287	3,184	41.6	24
2020	121	2,113	3,516	5,336	29.1	31
2021	549	6,989	29,629	40,869	54.0	54
2022	75	990	2,277	2,982	30.4	27
합계	832	11,390	39,704	55,483	47.7	30

자료: 저자 작성

우선 아프리오리 알고리즘을 통하여 기사와 댓글에서 고빈도 단어 집합을 추출하였다(<표 IV-9> 참조). 뉴스 기사에는 가상자산 혹은 암호화폐 소득 과세 관련 정보 전달을 위한 전문적 어휘가 빈번하게 등장한다. 물론 해당 단어집단의 높은 지지도는 특정 어휘를 이용하여 문장을 추출한 사실을 반영하고 있으므로 해석에 주의를 기울여야 한다. 한편 댓글에서는 ‘정부’, ‘국민’, ‘나라’, ‘민주당’, ‘정권’ 등 정치적 어휘가 빈번하게 등장하는 것이 특징적이며, 앞 절의 유튜브 댓글 분석에서 중요성이 드러난 ‘보호’ 키워드 역시 발견되었다.

〈표 IV-9〉 가상자산 과세 관련 네이버 뉴스 기사 및 댓글 고빈도 단어 집합

기사		댓글	
단어 집합	지지도	단어 집합	지지도
{‘가상’}	0.498	{‘세금’}	0.216
{‘자산’}	0.438	{‘정부’}	0.081
{‘과세’}	0.407	{‘코인’}	0.065
{‘화폐’}	0.403	{‘화폐’}	0.058
{‘가상’, ‘자산’}	0.374	{‘국민’}	0.046
{‘암호’}	0.288	{‘주식’}	0.042
{‘암호’, ‘화폐’}	0.279	{‘가상’}	0.042
{‘소득’}	0.203	{‘사람’}	0.041
{‘가상’, ‘과세’}	0.200	{‘나라’}	0.038
{‘가상’, ‘화폐’}	0.189	{‘과세’}	0.038
{‘과세’, ‘자산’}	0.175	{‘민주당’}	0.038
{‘거래’}	0.167	{‘보호’}	0.035
{‘가상’, ‘과세’, ‘자산’}	0.163	{‘비트코인’}	0.034
{‘정부’}	0.144	{‘정권’}	0.033
{‘과세’, ‘화폐’}	0.140	{‘가상’, ‘화폐’}	0.032
{‘세금’}	0.134	{‘인정’}	0.031
{‘거래소’}	0.117	{‘투자’}	0.030

자료: 분석 결과를 바탕으로 저자 작성

다음으로 연관규칙학습을 진행하였는데, 기사 및 댓글의 어휘 사용 밀도 차이를 반영하여 기사는 신뢰도 문턱값 0.9, 댓글은 신뢰도 문턱값 0.7을 설정하여 연관규칙을 추출하고, 이를 〈표 IV-10〉에 향상도 순으로 정리하였다. 앞 절에서 살펴본 바와 유사하게 ‘가상’, ‘화폐’와 ‘암호’, ‘화폐’ 등의 합성어가 각각 두 개의 단어로 인식되어 연관관계가 높은 것으로 나타났다. 댓글 분석 결과 ‘보호’가 포함된 문장은 3.5%(후향지지도)로 그 빈도가 매우 높지는 않지만, ‘투자자’와 ‘세금’이 같은 문장에 있을 때 ‘보호’가 등장할 확률(신뢰도)은 0.583으로 높아졌다. 이에 따라 {투자자, 세금} → {보호}의 연관규칙이 16.8로 높게 측정되었다. 유튜브 댓글에서와 유사하게 네이버 댓글 작성자들에게 투자자 보호가 중요한 개념으로 나타난다는 것을 확인할 수

있다. (가상, 인정) → (화폐)도 향상도 14.1을 보여 댓글 작성자들이 가상화폐를 ‘인정’할 것을 중요하게 생각함을 확인하였다. 이 또한 유튜브 댓글 분석과 유사한 결과로, 댓글의 경우 두 플랫폼 간 의견의 차이가 크지 않은 것으로 나타났다.

〈표 IV-10〉 가상자산 과세 관련 네이버 뉴스 기사 및 댓글의 연관규칙

구분	전항	후항	전항지지도	후항지지도	신뢰도	향상도
기사	{‘유예’, ‘자산’}	{‘가상’, ‘과세’}	0.032	0.200	0.937	4.674
	{‘정부’, ‘과세’, ‘암호’}	{‘화폐’}	0.022	0.403	0.996	2.474
	{‘거래’, ‘과세’, ‘암호’}	{‘화폐’}	0.022	0.403	0.996	2.474
	{‘가상’, ‘정부’, ‘암호’}	{‘화폐’}	0.020	0.403	0.996	2.473
	{‘가상’, ‘과세’, ‘암호’}	{‘화폐’}	0.034	0.403	0.995	2.471
	{‘가상’, ‘과세’, ‘암호’, ‘자산’}	{‘화폐’}	0.027	0.403	0.993	2.467
	{‘암호’, ‘부과’}	{‘화폐’}	0.024	0.403	0.993	2.466
	{‘과세’, ‘암호’, ‘소득’}	{‘화폐’}	0.027	0.403	0.990	2.460
	{‘내년’, ‘과세’, ‘암호’}	{‘화폐’}	0.022	0.403	0.988	2.453
	{‘암호’, ‘소득’}	{‘화폐’}	0.043	0.403	0.988	2.453
댓글	{‘투자자’, ‘세금’}	{‘보호’}	0.006	0.035	0.583	16.833
	{‘정부’, ‘화폐’}	{‘가상’}	0.007	0.042	0.622	14.800
	{‘암호’}	{‘화폐’}	0.013	0.058	0.844	14.472
	{‘가상’, ‘정부’}	{‘화폐’}	0.005	0.058	0.830	14.227
	{‘가상’, ‘인정’}	{‘화폐’}	0.004	0.058	0.821	14.074
	{‘가상’}	{‘화폐’}	0.042	0.058	0.760	13.030
	{‘화폐’}	{‘가상’}	0.058	0.042	0.548	13.030
	{‘화폐’, ‘세금’}	{‘가상’}	0.019	0.042	0.534	12.694
	{‘투자자’}	{‘보호’}	0.015	0.035	0.430	12.408
	{‘가상’, ‘세금’}	{‘화폐’}	0.014	0.058	0.710	12.171

자료: 분석 결과를 바탕으로 저자 작성

마지막으로 [그림 IV-2]는 아프리오리 알고리즘을 통해 생성한 지지도 0.02 이상의 뉴스 기사 단어 집합 중 크기가 2인 집합으로 구성한 의미망을 시각화한 그림이다. 기사의 경우 가상자산 투자 소득 과세 관련 정보 전달성

1) 설문조사 응답 분석

본 소절에서 분석하는 텍스트는 2022년 8월 29일부터 2022년 10월 4일 사이에 세 차례 진행된 설문에서 정부가 가상자산 발생 소득에 세금을 부과할 계획에 대하여 응답자가 긍정 혹은 부정적으로 생각하는 이유를 주관식 응답한 내용이다. 총 2,132개의 응답 중 지나치게 짧은 답변 등 분석이 불가능한 응답을 제외한 결과, 최종 분석 표본은 1,721개의 문장으로 이루어졌다.

〈표 IV-11〉은 설문 응답 텍스트에서 아프리오리 알고리즘을 통해 추출한 고빈도 단어 집합을 지지도 순으로 제시한다. 뉴스 기사 분석과 유사하게 가상자산 투자 소득 과세와 직접적인 관련성이 높은 어휘들이 집중적으로 등장한다. 설문 응답이 댓글보다는 기사에 가까운 정제된 언어로 이루어져 있음을 알 수 있다.

〈표 IV-11〉 가상자산 과세 관련 설문 응답의 고빈도 단어 집합

단어 집합	지지도	단어 집합	지지도
{‘소득’}	0.386	{‘투자’}	0.092
{‘과세’}	0.370	{‘과세’, ‘소득’, ‘가상’}	0.092
{‘가상’}	0.350	{‘자산’, ‘과세’, ‘소득’}	0.091
{‘자산’}	0.340	{‘세금’, ‘가상’}	0.088
{‘자산’, ‘가상’}	0.274	{‘자산’, ‘세금’}	0.086
{‘세금’}	0.271	{‘발생’}	0.084
{‘과세’, ‘소득’}	0.207	{‘자산’, ‘과세’, ‘소득’, ‘가상’}	0.083
{‘과세’, ‘가상’}	0.162	{‘화폐’, ‘가상’}	0.082
{‘자산’, ‘과세’}	0.159	{‘부과’, ‘세금’}	0.069
{‘소득’, ‘가상’}	0.143	{‘자산’, ‘가상’, ‘세금’}	0.068
{‘자산’, ‘소득’}	0.142	{‘거래’}	0.066
{‘자산’, ‘과세’, ‘가상’}	0.139	{‘주식’}	0.064
{‘자산’, ‘소득’, ‘가상’}	0.127	{‘발생’, ‘소득’}	0.063
{‘세금’, ‘소득’}	0.125	{‘부과’, ‘자산’}	0.051
{‘화폐’}	0.103	{‘부과’, ‘가상’}	0.049
{‘부과’}	0.098	{‘세율’}	0.048

자료: 분석 결과를 바탕으로 저자 작성

다음으로 <표 IV-12>는 기사 및 댓글에서 각각 신뢰도 0.5를 문턱값으로 추출된 연관규칙을 향상도 순으로 보여준다. {자산, 가상, 세율} → {적용}, {형평성, 가상} → {자산, 과세} 등 가상자산으로 얻은 소득에도 기타소득과 유사한 세율을 적용하는 것이 형평성에 맞다고 보는 의견과 관련한 연관규칙이 다수 나타났다. 투자자 보호 등을 강조하는 댓글의 의견과 뚜렷한 차이가 엿보인다.

<표 IV-12> 가상자산 과세 관련 설문 응답의 연관규칙

전항	후항	전항지지도	후항지지도	신뢰도	향상도
{'자산', '가상', '세율'}	{'적용'}	0.021	0.040	0.514	12.996
{'자산', '세율'}	{'적용'}	0.024	0.040	0.512	12.963
{'자산', '적용', '가상'}	{'세율'}	0.021	0.048	0.528	11.077
{'자산', '적용'}	{'세율'}	0.024	0.048	0.512	10.750
{'부과', '소득', '가상'}	{'자산', '세금'}	0.021	0.086	0.667	7.752
{'부과', '자산', '소득'}	{'세금', '가상'}	0.022	0.088	0.632	7.198
{'부과', '가상'}	{'자산', '세금'}	0.049	0.086	0.583	6.783
{'부과', '자산'}	{'세금', '가상'}	0.051	0.088	0.557	6.346
{'과세', '공정', '가상'}	{'자산', '소득'}	0.012	0.142	0.857	6.046
{'자산', '공정', '과세'}	{'소득', '가상'}	0.012	0.143	0.857	5.997
{'공정', '가상'}	{'자산', '과세', '소득'}	0.020	0.091	0.529	5.840
{'자산', '공정'}	{'과세', '소득', '가상'}	0.020	0.092	0.529	5.767
{'과세', '발생', '가상'}	{'자산', '소득'}	0.021	0.142	0.806	5.682
{'공정', '과세'}	{'자산', '소득', '가상'}	0.015	0.127	0.692	5.465
{'세금', '투자'}	{'부과'}	0.022	0.098	0.526	5.360
{'자산', '과세', '발생'}	{'소득', '가상'}	0.022	0.143	0.763	5.339
{'과세', '공정'}	{'자산', '소득'}	0.015	0.142	0.731	5.154
{'과세', '공정'}	{'소득', '가상'}	0.015	0.143	0.731	5.112
{'형평성', '가상'}	{'자산', '과세'}	0.013	0.159	0.783	4.916
{'발생', '가상'}	{'자산', '소득'}	0.037	0.142	0.688	4.849
{'자산', '세금', '과세'}	{'소득', '가상'}	0.023	0.143	0.692	4.843

자료: 분석 결과를 바탕으로 저자 작성

[그림 IV-4]는 아프리오리 알고리즘을 통해 생성한 지지도 0.01 이상의 설문 응답 단어 집합 중, 크기가 2인 집합의 의미망을 시각화한 그림이다. 고

지만, 댓글의 경우 가장 지지도가 높은 ‘세금’도 0.067로 낮은 지지도를 보여 사용 어휘가 산발적으로 분포함을 알 수 있다. 앞선 뉴스 기사 분석의 경우와는 달리 설문 응답의 경우 특정 어휘를 이용하여 추출한 문장만을 대상으로 한 분석이 아니므로, 해당 지지도 차이는 두 텍스트의 어휘 사용 밀도의 차이를 정확히 반영하고 있다. 특히 댓글의 경우 정치 관련 어휘가 많아 가상자산 과세에 대한 의견을 파악하기 어렵지만, 설문 응답의 경우 가상자산 과세 관련 어휘가 많이 사용되어 정제된 의견을 파악하기 쉬운 것으로 나타났다.

〈표 IV-13〉 2022년 네이버 댓글과 설문 응답의 크기 1인 고빈도 단어 집합

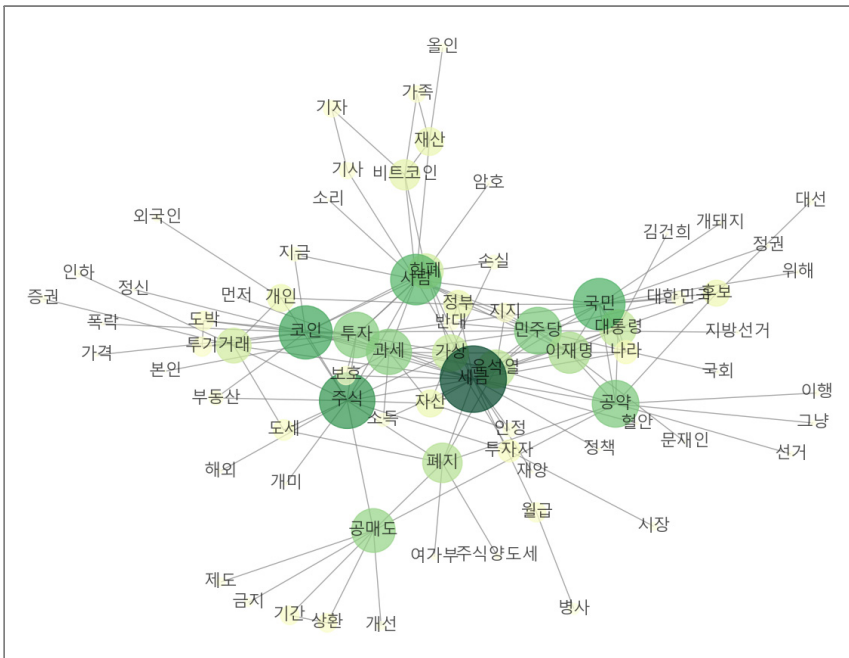
2022년 네이버 댓글		설문 응답	
단어 집합	지지도	단어 집합	지지도
{‘세금’}	0.067	{‘소득’}	0.386
{‘코인’}	0.065	{‘과세’}	0.370
{‘사람’}	0.054	{‘가상’}	0.350
{‘주식’}	0.045	{‘자산’}	0.340
{‘공약’}	0.042	{‘세금’}	0.271
{‘국민’}	0.042	{‘화폐’}	0.103
{‘민주당’}	0.041	{‘부과’}	0.098
{‘이재명’}	0.037	{‘투자’}	0.092
{‘과세’}	0.036	{‘발생’}	0.084
{‘비트코인’}	0.033	{‘거래’}	0.066
{‘대통령’}	0.033	{‘주식’}	0.064
{‘윤석열’}	0.031	{‘세율’}	0.048
{‘가상’}	0.031	{‘수익’}	0.046
{‘나라’}	0.031	{‘투기’}	0.046
{‘폐지’}	0.030	{‘정부’}	0.044
{‘투자’}	0.028	{‘모든’}	0.044

자료: 분석 결과를 바탕으로 저자 작성

〈그림 IV-5〉는 아프리ורי 알고리즘을 통해 생성한 지지도 0.02 이상의 네이버 댓글 단어 집합 중, 크기가 2인 집합을 모아 네트워크로 시각화한 그

림을 보여준다. [그림 IV-4]의 설문 응답 의미망과 비교하였을 때, 관련 어휘의 밀집도가 떨어지며 정치, 비트코인 투자, 공매도 등과 관련한 개별 클러스터들이 발견된다. 정치 관련 클러스터는 특히 2022년에 시행된 대통령 선거의 영향 등으로 더욱 뚜렷하게 나타나는 것으로 보이며, 이로 인해 가상자산 소득 과세와 직접적으로 관련이 있는 의미 구조를 파악하는 데에 어려움이 발생한다.

[그림 IV-5] 가상자산 과세 관련 2022년 네이버 뉴스 기사 댓글의 의미망 시각화



자료: 분석 결과를 바탕으로 저자 작성

결론적으로 설문조사 응답의 정제된 의견과 비교할 때, 댓글의 텍스트는 산만하고 정치 관련 언급의 비중이 높아 가상화폐 관련 직접적인 의견을 파악하는 데 쉽지 않은 것으로 보인다. 하지만 공매도나 코인 투자 관련 클러스터가 발생하는 등 가상화폐 자산 과세에 관심을 가지는 네티즌이 어떠한 세부 주제에 관심을 가지는지 분석하는 경우 유용하게 사용될 수 있을 것으

로 보인다. 즉 특정 주제로 정제되지 않은 의견이 더욱 많이 포함된 만큼, 잡음을 세밀하게 제거하는 과정을 거칠 때 유의미한 분석을 이끌어낼 수 있을 것으로 보인다.

V. 인터넷 미디어 여론의 토픽 모델링 및 감성분석

본 장에서는 인터넷 미디어 여론과 설문조사에서 얻어진 텍스트를 정량적으로 분석하여 비교한다. 우선 제V장 제1절에서는 해당 텍스트들을 잠재 디리클레 할당(Latent Dirichlet Allocation, LDA) 토픽 모델링 방법론을 통해 여러 주제로 분류한 후, 분류된 주제들이 가지고 있는 내용과 구성비를 비교한다. 제V장 제2절에서는 설문 텍스트를 통해 조세정책에 대한 긍·부정의 감성을 도출할 수 있는 감성분류기를 만들어 인터넷 미디어 여론의 감성을 추정해 보고, 이것이 설문의 긍·부정 응답 비율에 비해 얼마나 편향적인지 확인해 보고자 한다.

1. 토픽 모델링 분석

본 절에서는 잠재 디리클레 할당(Latent Dirichlet Allocation, LDA)을 활용한 토픽 모델링을 통해 뉴스에 대한 댓글과 설문조사에서 얻어진 텍스트에서의 주제를 추출해 보고자 한다. 단순히 뉴스 기사를 게재하고 그에 대한 독자들의 댓글을 기록하는 인터넷 미디어에서 조세정책 관련 키워드(keyword)를 기준으로 기사를 수집(crawling)하였다고 하여 반드시 조세정책 텍스트 자료만 수집되는 것은 아니기 때문에, 관련 높은 텍스트를 선별하여 수집하는 것 외에도 수집된 텍스트가 적절한 주제로 이루어져 있는지 사후에 점검하는 것이 바람직하다. 만약 수집된 텍스트에 관심 주제에 적절하지 않은 자료가 혼동된 경우 이는 잡음(noise)에 해당하며, 해당 자료로 학습한 감성분류기는 편의(bias)를 내포할 우려가 있다. 이러한 점을 고려하면 수집된 텍스트에서 공통적으로 나타나는 가능한 주제를 추출해 보고, 관련 없는 주제에 해당하는 텍스트가 수집된 경우 수집과정을 다시 거치는 것이 좋다.

우선 가상자산에 기타소득세에 대한 과세를 주제로 하는 인터넷 뉴스 및 댓글을 대상으로 LDA 분석을 수행하였다. 수집된 자료는 대표적 플랫폼 서비스인 네이버에서 ‘과세’를 반드시 포함해야 할 키워드로 하고 ‘암호화폐, 가상화폐, 가상자산’을 관련 키워드로, ‘규제, 종부세’는 제외해야 할 키워드로 하여 뉴스 기사를 검색하고, 검색된 기사와 해당 기사에 독자들이 작성한 댓글을 수집하였다.

암호화폐 과세에 관련된 기사는 2018년 이후부터 직접 연관된 기사가 검색되므로 기간은 2018년 1월~2022년 6월로 하였다. 본고에서는 편의상 연도별로 주제를 점검하였다. <표 V-1>을 보면 암호화폐 과세와 관련된 기사는 2021년까지 지속 증가하다가 2022년 들어 다시 감소하고 있다. 이와 관련한 평균 댓글 수는 암호화폐 과세 관련 기사가 본격적으로 등장하기 시작한 2018년이 62.3개로 가장 많았으며, 2019년과 2020년 들어 각각 42.6개, 28.8개로 점차 감소하다가 2021년 들어 다시 53.9개로 급증하였다. 기사에 대한 평균 댓글 수가 기사에 대한 독자들의 관심을 반영한다고 가정하면, 암호화폐 과세 관련 기사가 등장하기 시작한 2018년에 관심이 가장 높았고, 이후 과세에 대한 논의가 지연되었던 2019~2020년 동안 관심이 낮았다가 2021년 다시 주목받기 시작한 것으로 해석할 수 있다.

<표 V-1> 암호화폐 과세 관련 기사 및 댓글 자료 건수

(단위: 건 개)

구분	2018	2019	2020	2021	2022
기사 수	32	54	120	545	75
댓글 수	1,995	2,287	3,516	29,629	2,277
평균 댓글 수	62.3	42.6	28.8	53.9	30.4
언론사 수	14	24	31	54	27

자료: 저자 작성

LDA(Latent Dirichlet Allocation)는 문서에 사용된 단어를 기반으로 예상되는 주제에 대한 확률을 계산하는 과정이다. 문서 집합 전체 문서 개수를 D , 사전 정의된 토픽의 수 K , d 번째 문서의 총 단어 개수 N , d 번째 문서의

n 번째 단어 $w_{d,n}$, 디리클레 분포의 모수 α, β 가 주어졌을 때, ϕ_k 를 각 단어가 k 번째 토픽에 속할 확률벡터라고 하고, θ_d 를 d 번째 문서가 가진 토픽 비중을 나타내는 벡터라고 하면, LDA에서는 전체 문서로 이루어진 말뭉치 내 한 단어가 쓰일 조건부확률 $P(w_{1:D} | \phi_{1:K}, \theta_{1:D}, z_{1:D}; \alpha, \beta)$ 이 ① 각 토픽의 말뭉치 내 단어에 대한 분포 $\prod_{i=1}^K P(\phi_i | \beta)$ 와 ② 각 문서가 지닌 토픽 분포 $\prod_{i=1}^D P(\theta_d | \alpha)$ 의 결합분포로 나타난다. 따라서 LDA는 문서의 한 단어가 관찰되었을 경우 ①과 ②를 추정하는 과정이다.⁴⁰⁾

$$P(\phi_{1:K}, \theta_{1:D}, z_{1:D} | w_{1:D}) = \frac{P(\phi_{1:K}, \theta_{1:D}, z_{1:D}, w_{1:D})}{P(w_{1:D})}$$

$$= \prod_{i=1}^K P(\phi_i | \beta) \prod_{i=1}^D P(\theta_d | \alpha) \left\{ \prod_{i=1}^N P(z_{d,n} | \theta_d) P(w_{d,n} | \phi_{1:K}, z_{d,n}) \right\}$$

한편 토픽의 개수는 임의로 정하기보다 이하와 같이 통일성 점수(coherence score)⁴¹⁾를 통해 가장 높은 점수를 부여받은 토픽 개수를 기준으로 한다. 여기에서 w_i 와 w_j 는 말뭉치 내 단어이며, 통일성 점수가 높을수록 바람직한 토픽의 개수라고 평가할 수 있다.

$$C(W) = \left\{ \sum_{w_i \in W} \left(\frac{\log \frac{P(w_i, w_j) + \epsilon}{P(w_i) P(w_j)}}{-\log(P(w_i, w_j) + \epsilon)} \right) \right\}_{j=1, \dots, N}$$

40) Gibbs 추출과 베이저인 추정을 통해 최적화하는 과정으로 추정한다.

41) Röder, M., Both, A., and Hinneburg, A., "Exploring the Space of Topic Coherence Measures," WSDM '15: Proceedings of the Eighth ACM International Conference on Web Search and Data Mining, February 2015, pp. 399~408.

가. 뉴스 기사와 댓글의 토픽 모델링

〈표 V-2〉는 LDA 추정을 통해 연도별 기사에서 추출한 토픽이며 추정 결과는 키워드로 구성되어 있으나, 키워드를 읽는 것만으로도 어떠한 토픽인지 짐작할 수 있을 정도로 연관성 있는 단어로 구성되어 있음을 알 수 있다.⁴²⁾ 2018년에는 기사의 수도 적었지만 토픽도 단지 암호화폐의 과세에 국한되지 않고 부동산, 주식 등 보유자산 수익에 대한 포괄적 과세 논의의 일환이었음을 알 수 있다. 2019년에는 가상화폐거래소인 ‘빗썸’과 같은 거래소에 과세하는 방안이 국회 등에서 논의되고 있었음을 알 수 있다. 또한 암호화폐 거래에 대한 과세를 거래소 자료를 통해 원천징수하는 방안 역시 논의하였다는 점을 짐작할 수 있다. 2020년 이후 토픽도 다양해지는 동시에 과세에 대한 논의도 구체적으로 진행되었음을 알 수 있다. 이 시기에는 암호화폐 거래에 대한 과세 근거의 토픽과 비거주자가 국내 거래소에서 거래할 경우 우리나라 세법의 적용 문제, 기술적 제약 등으로 세원을 정확히 파악할 수 없을 때 신고방식을 도입할 것인지 등에 대한 논의와 협의가 진행되고 있었다. 2021년에도 2020년과 같은 논의가 이어지고 있었으나, 선거를 앞두고 일시적으로 과세를 유예하였다는 점을 토픽을 통해 알 수 있다. 2022년에는 암호화폐 과세에 대한 다방면의 논의가 지속되고 선거로 인해 유예되었던 과세가 새 정부가 들어서며 재검토 중인 사실을 토픽을 통해 알 수 있다.

42) 본문의 표에서는 지면을 아끼기 위해 주제별 단어의 수를 10개 이하로 제한하였다. 한편 〈부록 4〉에서는 말뭉치(corpus)를 주제별로 묶어 단어의 빈도를 도식화하는 워드 클라우드(word cloud)로 텍스트 자료를 시각화하여 주제별 어휘들의 빈도나 등장 어휘들을 더 자세히 확인할 수 있도록 하였다.

〈표 V-2〉 가상자산소득 과세 관련 기사 토픽 모델 분석 결과

연도	키워드	기사 수	기사 비율
2018	소득, 통화, 주식, 수익, 투자자, 부과, 소득세	13	0.41
	부총리, 기획, 문제, 국회, 정보, 정책, 재정부, 논의	10	0.31
	블록체인, 기술, 시장, 기업, 비트코인, 부동산, 보유, 자금	9	0.28
	합계	(32)	(1.00)
2019	빚쟁, 외국인, 원천, 징수, 업계, 고객, 이익	18	0.33
	통화, 가격, 시장, 한국, 보유	17	0.32
	국회, 내년, 방침, 통과, 기재부, 개정, 뉴스, 정보	19	0.35
	합계	(54)	(1.00)
2020	거래세, 블록체인, 세제, 차익, 정책, 방식, 조세, 의견, 협회	24	0.20
	빚쟁, 국세청, 근거, 소득세법, 징수, 원천, 대상, 비거주자	20	0.17
	투자, 투자자, 시장, 디지털, 주식, 서비스, 발표	23	0.19
	금액, 분류, 취득, 내역, 포함, 가액, 검토, 성격, 차익, 방침	27	0.22
	사업자, 신고, 업계, 시행, 법률, 자금, 구축, 확인, 의무, 계좌	26	0.22
	합계	(120)	(1.00)
2021	뉴스, 경제, 보호, 위원장, 제도, 한국, 인정, 법	114	0.21
	주식, 공제, 기타, 기준, 적용, 차익, 소득세, 개정안, 경우	99	0.18
	비트코인, 코인, 가격, 신고, 정보, 사업자, 달러, 국세청	154	0.28
	의원, 민주당, 유예, 부총리, 정책, 입장, 국민, 후보, 논의	178	0.33
	합계	(545)	(1.00)
2022	조사, 대선, 대표, 여론, 이재명, 부동산, 뉴스, 세대, 유예, 선거	11	0.14
	소득, 공제, 소득세, 국회, 상자, 적용, 예정, 법, 시행	14	0.19
	비트코인, 달러, 가격, 시세, 미국, 올해, 전망, 증여, 평가	14	0.19
	블록체인, 사업, 메타버스, 문제, 플랫폼, 기술, 작품	14	0.19
	정부, 보호, 국민, 윤석열, 공개, 한도, 검토, 허용, 마련	22	0.29
	합계	(75)	(1.00)

주: 키워드는 10개 이하로 제한하였으며 () 안의 숫자는 합
 자료: LDA 토픽 모델 분석을 바탕으로 저자 작성

한편 주제의 분류에서 볼 수 있듯이, 뉴스 기사들의 경우에는 가상자산
 기타소득세에 대한 납세자들의 인식보다는 국회나 기획재정부에서 조세정책
 관련 법률을 검토하는 사실을 전달하거나, 가상자산의 기술적 근간이 되는
 블록체인 기술을 소개하거나, 대권주자들의 공약을 전달하는 주제가 많은

것을 확인할 수 있다. 예를 들어 2018년에는 28%의 주제가 블록체인 기술에 관련된 것이었으며, 2022년에는 44%(1주제+5주제)가 대권주자와 관련이 있는 주제였다.

〈표 V-3〉은 LDA 추정을 통해 연도별 기사의 댓글에서 추출한 토픽이다. 추정 결과는 키워드로만 구성되어 있지만, 대체로 조세정책 관련 정부 정책에 대한 과격한 불만을 토로하는 내용일 것으로 짐작할 수 있다. 중요한 것은 네이버 뉴스 수집과정에서 종부세 관련 기사들은 분명히 제외했음에도 불구하고 부동산 관련 과세에 대한 댓글이 상당히 많다는 것이다. 예를 들어 2018년에는 암호화폐 거래소에 대한 규제에 폐쇄될 수 있다는 내용의 기사와 그에 대한 댓글이 있었다. 그러나 여전히 부동산 관련 과세에 관한 내용이 대부분이라는 것을 해당 주제에 대한 댓글의 비율(44%)로부터 알 수 있다. 2019년 댓글도 34%는 전반적으로 증세 기조를 보이고 있는 정부에 대한 불만을 주제로 하고 있었다. 2022년도 30%의 댓글은 금융투자소득세와 공매도에 관한 내용을 주제로 하고 있었으며, 27%는 대선 후보들에 관한 내용을 주제로 하고 있었다. 이러한 점은 인터넷 미디어 여론에서 얻어지는 텍스트는 특정 조세정책에 대한 인식만 포함하는 것이 아니라, 다른 조세정책에 대한 의견이나 정권에 대한 평가 등 여러 가지 잡음(noise)으로 생각할 수 있는 주제들이 많이 섞여 있다는 것을 알 수 있다.

댓글의 주제들을 연도별로 살펴보면, 2019년에는 당시 정권에 대한 과격한 부정적인 의견들이 다수를 차지하였다. 그리고 본격적으로 암호화폐 거래에 대한 과세에 의견을 표출하는 댓글이 다수 등장한 것을 볼 수 있다. 2020년은 가상자산소득 과세에 저항하는 댓글이 많이 증가하였고, 그 내용도 당시 대통령과 정부에 대한 비난이 주를 이루고 있음을 알 수 있다. 이는 2021년에도 이어졌으나 댓글의 수는 약 3만건으로 폭증하는 모습을 볼 수 있다. 또한 가상자산소득 과세정책을 발의하는 당시 집권당을 선거에서 심판하겠다는 댓글도 상당수 있음을 알 수 있다. 2022년에는 선거 이후 가상자산소득 과세에 대한 정부의 입장이 재검토됨에 따라 댓글 수도 많이 감소하고 내용에도 다소 변화가 있음을 알 수 있다.

〈표 V-3〉 가상자산소득 과세 관련 댓글 토픽 모델 분석 결과

연도	키워드	댓글 수	댓글 비율
2018	나라, 투자, 말, 국민, 부동산, 돈	867	0.43
	화폐, 돈, 가상, 코인, 사람, 과세, 주식, 거래소, 규제, 우리	572	0.29
	세금, 정부, 생각, 인정, 서민, 거래소, 폐쇄	552	0.28
	합계	(1,991)	(1.00)
2019	주식, 나라, 니겟, 정권, 생각, 투자, 경제, 시장	801	0.35
	세금, 돈, 정부, 사람, 국민, 문, 재앙, 부과, 개	784	0.34
	세금, 화폐, 인정, 비트코인, 과세, 가상, 소득, 코인	697	0.31
	합계	(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
	생각, 부동산, 정책, 투자, 비트코인, 경제, 국가	602	0.17
	화폐, 가상, 거래소, 코인, 인정, 자산, 암호	751	0.22
	정부, 국민, 정권, 사람, 세금, 개, 소득	624	0.18
	합계	(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
	사람, 비트코인, 코인, 부동산, 투기, 시장, 도박	4,938	0.17
	세금, 화폐, 가상, 보호, 정부, 자산, 인정, 국가	5,355	0.18
	코인, 주식, 과세, 투자, 거래소, 소득	4,562	0.16
	합계	(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
	합계	(2,160)	(1.00)

주: 키워드는 10개 이하로 제한하였으며 () 안의 숫자는 합
 자료: LDA 토픽 모델 분석을 바탕으로 저자 작성

나. 조세정책 관련 설문의 토픽 모델링

〈표 V-4〉는 3회차에 걸쳐서 이루어진 가상자산 기타소득세 관련 개방형 설문에서 긍정과 부정 의견에 대한 응답의 주제를 LDA로 분석한 결과이다. 응답 중 의견의 근거가 될 수 없는 불성실한 답변을 제외⁴³⁾한 수정표본에서

43) 예를 들어 ‘그냥’, ‘모르겠음’ 등의 응답이 제외되었다.

주제를 추출하였으며, 긍정과 부정 응답에서 모두 7개의 주제가 추출되었다. 긍정 답변의 경우 주제별로 응답을 분류하였을 때 각 토픽에 해당하는 응답이 12~20%의 비율로 고르게 분포된 것으로 나타나며, 부정의 경우는 각 토픽이 8~20%의 비중을 차지하는 것으로 나타났다.

긍정 답변의 경우 20%에 해당하는 비율로 나타난 ‘화폐, 돈, 수익, 사람, 가치, 이득, 자금’ 키워드를 지닌 토픽은 수익에 대해 당연히 과세해야 한다는 논리의 주제로 해석할 수 있다. 두 번째 주제는 ‘무분별, 투자, 규제, 투기, 범죄, 방지, 조세, 제도’ 키워드로 구성되어 있으며, 무분별한 투자를 규제하고 투기와 범죄를 방지하는 목적의 조세제도로 가상자산소득 과세를 긍정적으로 생각하는 응답임을 알 수 있다. 세 번째 주제는 ‘불로소득, 일반, 개인, 소득세, 세율, 적용, 원칙, 동일’ 키워드로, 가상자산 투자에 의한 소득이 불로소득이며 과세율 20%는 소득세율에 비추어볼 때 과하지 않다는 입장인 것을 알 수 있다. 네 번째 주제는 ‘주식, 재산, 경우, 세금, 부과, 해당, 이유, 부여, 금액’의 키워드로 구성되어 주식거래의 양도세에 비추어볼 때 가상자산소득 과세는 형평성에 적합하다는 입장인 것으로 볼 수 있다. 다섯 번째부터 일곱 번째 주제까지는 키워드만으로 뚜렷한 주제를 특정 짓기 어렵다. 그러나 가상자산소득 과세에 대해 필요하며 공평하다고 생각하는 응답임을 알 수 있다.

부정 응답의 주제도 긍정 응답과 많은 키워드를 공유하고 있으나 다른 논리에 의해 반대 의견이 서술되고 있다. 첫 번째 주제는 ‘위험, 투자, 돈, 수단, 과세, 부당, 생각, 가치, 때문, 정도’ 키워드로 구성되어 있으며, 위험을 감수한 투자에 따르는 수익에 과세하는 것은 부당하다는 주제임을 알 수 있다. 두 번째 주제 키워드는 ‘가상, 자산, 실물, 시장, 실제 현금, 변동, 소득세’로, 현금과 같은 실물 자산과 달리 변동성이 큰 자산으로 암호화폐(가상 자산)에 과세하는 것이 부당하다는 주제의 응답임을 알 수 있다. 세 번째 주제는 ‘가상, 화폐, 법, 인정, 손해, 안정’으로 암호화폐가 법 제도적으로 자산으로 인정되어야 과세가 정당하다는 주제의 의견이다. 네 번째 주제는 ‘투자자, 손실, 보호, 제도, 과세, 정부, 규제, 정책’으로 시스템 리스크에 의한 투

자산손실을 보호하는 제도적 장치 미비에도 불구하고 정부가 지나치게 규제 위주의 정책을 펼친다는 주제의 응답일 가능성이 높다. 다섯 번째 주제는 ‘세금, 부과, 과도, 국가, 수익, 부분’으로 20%의 세율이 과하다는 주제의 응답일 수 있다. 여섯 번째와 일곱 번째 주제는 특정하기 어려운 키워드로 구성되어 있다.

이상에서 확인할 수 있듯이 가상자산 기타소득세 관련 개방형 설문에서는 해당 정책에 밀접한 연관이 있는 주제들이 절대적인 다수를 차지하는 것을 확인할 수 있다. 이를 바탕으로 볼 때 조세정책에 대해서 개방형 질문에 대해서 응답한 텍스트는 조세정책에 대한 선호를 잘 반영하고 있을 것으로 판단한다. 이후 감성사전 구축 작업에서는 설문조사에서 얻은 텍스트를 바탕으로 구축을 진행하였다.

〈표 V-4〉 조세정책 관련 개방형 설문에 따른 긍정·부정 응답 토픽 분석

긍/부정	키워드	관련 응답 수	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
	합계	(1,364)	(1.00)
부정 응답	위험, 투자, 돈, 수단, 과세, 부당, 생각, 가치, 때문, 정도	126	0.20
	가상, 자산, 실물, 시장, 실제 현금, 변동, 소득세	83	0.14
	가상, 화폐, 법, 인정, 손해, 안정	84	0.14
	투자자, 손실, 보호, 제도, 과세, 정부, 규제, 정책	73	0.12
	세금, 부과, 과도, 국가, 수익, 부분	120	0.19
	세율, 때문, 부정, 적용, 개인, 부담, 이익, 발생	82	0.13
	소득, 자산, 가상, 주식, 거래, 금융, 경우, 기준, 형평	49	0.08
	합계	(617)	(1.00)

주: 키워드는 최대 10개로 제한하였으며 () 안의 숫자는 합

자료: LDA 토픽 모델 분석을 바탕으로 저자 작성

2. 조세 수용성을 측정할 감성사전의 구축

머신러닝 기법의 발달과 광범위한 적용으로 텍스트 데이터를 분류하고 분석하는 기법 또한 비약적인 발전을 이루었다. 텍스트는 생산되는 분량이 방대할 뿐만 아니라 텍스트의 작성과 동시에 그 의미가 즉시 해석될 수 있는 특성으로 인해 텍스트를 활용하여 투표 결과를 예측⁴⁴⁾하거나 시장의 기대를 측정하는 데 널리 사용되고 있다(김수현 외, 2020). 또한 특정 정책에 대한 반응을 점검하고 예측하는 데에도 텍스트는 주요한 분석 자료로 활용될 수 있다. 익명이 보장되는 글에서 글쓰이는 굳이 자기 생각을 숨기거나 왜곡할 이유가 없으므로, 분석 목적에 따라 적합한 텍스트를 활용할 경우 매우 효율적인 분석과 예측을 할 수 있다.

가. 감성사전의 정의와 구축 방법

시장의 기대를 측정하고 정책에 대한 반응 등을 점검하는 데는 감성사전(sentiment lexicon)을 필요로 한다. 감성사전은 사전 학습된 모형을 통해 특정 단어 또는 형태소의 조합이 어떠한 감성을 띄는지 측정함으로써 텍스트 전체의 감성을 정량적으로 측정할 수 있는 방법이다. 감성사전을 구축하기 위해 우선 모형을 훈련할 수 있는 자료가 필요한데, 해당 자료는 특정 텍스트가 지닌 감성이 표기되어 있어야 한다. 따라서 감성사전을 구축하는 데는 분류된 자료⁴⁵⁾를 얻는 과정이 필요한데, 이 과정에서 기존 텍스트를 사람이 일일이 분류하거나 텍스트의 감성을 분류할 수 있는 기준이 되는 지표를 찾아냄으로써 기계적으로 분류하기도 한다. 따라서 학습자료를 얻는 과정이 감성사전을 구축하는 과정에서 가장 큰 비용이 소요된다.

학습자료가 준비되면 감성사전의 모형을 선택해야 한다. 감성사전 모형은 분류와 예측 성능이 주요한 기능이므로 보편적으로 머신러닝을 활용한다. 분류에 활용되는 머신러닝 모형은 로지스틱 회귀, 서포트 벡터 머신(SVM,

44) Five Thirty Eight사의 경우 트위터를 활용하여 미국 대선 결과를 정확히 예측하기도 하였다.

45) 이 학습자료가 감성사전의 감성분류 성능을 좌우함은 말할 것도 없다.

Support Vector Machine), 나이브 베이즈 분류기(naive Bayes classifier)와 같이 분류에 특화된 모형을 사용하거나 비선형 모형인 각종 인공신경망(neural networks)을 사용한다. 인공신경망은 그 성능과 광범위한 활용 가능성으로 최근 들어 적용 범위를 더욱 넓혀 가고 있다.

나. 조세 수용성을 측정할 수 있는 감성사전의 구축

본 연구에서는 개방형 주관식 설문조사를 통해 긍정과 부정 의견이 포함된 텍스트 자료를 구성하였다. 3회에 걸친 설문조사에서 “가상자산 소득 과세 정부는 가상자산에서 발생한 소득에 대해 20%의 세율을 적용하여 과세하는 것을 검토하고 있습니다. 귀하는 가상자산 소득에 대해서 과세하는 것에 대해 어떠한 입장이십니까?”라는 질문에 대해 긍정과 부정으로 답하게 되어 있으며, “가상자산 소득 과세에 긍정적으로 생각하는 이유는 무엇입니까?”라는 질문에 대해 긍정으로 답변한 피설문자의 이유를, “가상자산 소득 과세에 부정적으로 생각하는 이유는 무엇입니까?”라는 질문에 대해서는 부정으로 답변한 피설문자의 이유를 문장으로 답변하도록 되어 있다. 설문 결과 1회차에서 총 715개(긍정: 479개, 부정: 236개), 2회차에서 총 709개(긍정: 501개, 부정: 208개), 3회차에서 총 708개(긍정: 497개, 부정: 211개)의 답변을 표본으로 구하였다. 3회에 걸친 설문이 독립적으로 추출된 표본에 대해 이루어졌음에도 불구하고, 긍정 답변의 비율은 각각 67.0%, 70.7%, 70.2%으로, 긍정 비율이 70% 내외로 매우 일정한 수준이다.⁴⁶⁾ 이러한 긍부정 비율의 일관성(consistency)으로 유추해 볼 때, 각각의 설문조사는 가상자산 소득 과세에 대한 민간의 인식을 잘 반영하고 있다고 판단할 수 있다. 또한 긍부정의 이유로 제시한 문장들도 구체적인 이유를 제시한 답변이 많아 가상자산 소득 과세에 대한 감성을 측정할 감성분류기를 학습하기에 적합하다. 이하에서는 설문문의 결과를 토대로 가상자산 소득 과세에 대한 감성을 분류할 수 있는 학습기를 학습하고 검증하는 과정에 관해 기술한다.

46) 평균은 69.3%이다.

1) 감성분류기: 트랜스포머 모형

머신러닝과 자연어 분석의 발달로 감성분석기를 구축하는 방법은 매우 다양하다. 본 연구에서는 가장 최근에 개발되어 자연어 처리뿐만 아니라 이미지 등 다양한 분야에서 월등한 성능을 자랑하는 트랜스포머(transformer) 모형으로 감성분류기를 학습하였다. 트랜스포머(Vaswani et. al., 2017)는 2017년 구글(Google)이 제안한 seq2seq 딥러닝(sequence-to-sequence deep learning) 모형으로 기계번역을 위해 개발된 거대 인공지능 모형이다. 기계번역은 어떤 언어(source language)의 단어 배열을 다른 언어(target sequence)의 단어 배열로 변환하기 위한 모형이나, 서로 다른 언어 간 기계번역에서 기존 모형으로 난해했던 부분인 서로 다른 배열의 길이(예: 한글 5개 단어를 번역하면 영어로 7개 단어가 되는 경우) 문제에서 벗어날 뿐만 아니라, 속성이 다른 자료 간 변환(예: 문장 \Rightarrow 이미지)도 가능하다는 특징으로 모형의 이름(트랜스포머)이 그 기능을 잘 설명한다. 자료의 변환이 가능한 이유는 트랜스포머가 입력자료의 특성을 압축하는 인코더(encoder)와 압축된 특성을 해제하며 원하는 자료의 형태로 출력하는 디코더(decoder)로 구성되어 있기 때문이다.

트랜스포머의 학습과정은 입력자료의 압축(encoding)과 해제(decoding)과정에서 정답에 해당하는 배열의 확률을 높이는 과정으로 이루어진다. 일반적으로 트랜스포머는 압축과 해체를 통해 입력자료와 다른 형태(예: 다른 언어, 이미지 등)를 출력해 내지만, 감성사전을 구축하기 위해 필요한 출력은 '긍정, 부정'이다. 따라서 감성분류기 구축을 위해 학습할 디코더 부분은 다른 언어모형에 비해 간단한 구조를 지닌다. 트랜스포머가 기존의 딥러닝 모형과 차별되는 특징은 집중(어텐션, attention) 알고리즘에 있다. 이는 다시 자기집중(self-attention)으로 입력자료 중에 정보량이 많은 자료에 대한 입력가중치를 높이는 방식으로 학습이 이루어진다는 의미이다. 기존 언어모형에서 활용되던 재귀신경망(recurrent neural network, RNN)에서는 단어의 입력순서에 따라 정보량의 가중치가 달라지는 한계가 있었으며, 합성곱신경망(convolutional neural network, CNN)에서는 사전 정의한 초모수(hyperparameter)

인 필터의 길이 이상의 문장 내에 포함된 단어 간 관계를 읽어낼 수 없다는 한계가 있었다. 따라서 재귀신경망을 사용할 경우 문장에서 가장 중요한 단어가 앞쪽에 나타날 때 모형의 성능 저하로 이어질 수 있으며, 합성곱신경망의 경우 문장이 길어질 때 모형이 문장의 문맥을 분석해 내지 못할 가능성이 높다. 자기집중 알고리즘으로 학습하는 트랜스포머 모형의 경우, 전체 문장에서 밀접한 관계를 지닌 단어 조합의 모든 경우의 수에 대해 정보량을 측정하여 높은 정보량에 해당하는 조합에 집중함으로써 모형의 성능을 개선한다. 감성분류기에서도 단어 간 관계를 중점적으로 고려하므로 분류 성능을 높일 수 있다. 예를 들어 “가상자산 소득 과세에 긍정적으로 생각하는 이유는 무엇입니까?”라는 질문에 대해 “가상자산 투자는 안전하지 않으므로 위험한 투기행위를 방지하기 위해”라고 답변한 문장과 “가상자산 소득 과세에 긍정적으로 생각하는 이유는 무엇입니까?”라는 질문에 대해 “가상자산은 개인이 위험을 감수하고 투자하므로 정부가 가상자산 거래시스템의 안전을 보장하지 않는 이상 과세하는 것은 이치에 맞지 않음”이라고 부정의 이유를 답변한 경우가 있다고 가정하자. 두 답변 모두 ‘가상자산, 위험, 안전’이라는 단어의 조합이 들어 있음을 알 수 있다. 그런데도 첫 번째 답변은 ‘긍정’으로 분류되어야 하며, 두 번째 답변은 ‘부정’으로 분류되어야 한다. 따라서 같은 단어 조합이 있는 서로 다른 문장을 각각의 감성으로 분류해 내기 위해 ‘가상자산, 위험, 안전’과 여타 단어들의 조합을 이룸으로써 각각 긍정과 부정의 감성으로 분류하도록 학습되어야 한다.

본 연구에서 구축한 감성분류기는 구글이 Vaswani et. al.(2017)의 이론을 기초로 개발하여 배포한 BERT(Bidirectional Encoder Representations from Transformers)를 기반으로 한다. BERT(Devlin et al., 2018)는 사전 분류되지 않은 위키피디아(Wikipedia)의 25억개 단어와 BookCorpus⁴⁷⁾의 8억개 단어로 사전 학습(pre-trained)된 언어모형이다. 사전 학습된 모형이므로 사전 분류(labeling)된 학습자료로 추가학습(fine-tuning)을 거칠 경우 높은 성능을

47) BookCorpus는 무료 배포된 1만 1,038권 소설의 약 7,400만 문장으로 구성된 거대 텍스트 자료이다.

나타내는 것으로 알려져 있다. BERT의 기본구조는 트랜스포머(12개 또는 24개)를 직렬층(sequential layers)으로 중층 구성하였으며, 단어 간의 관계인 문맥(context)을 인지하는 트랜스포머 모형의 특성이 잘 구현되어 있다. 그러나 감성분석은 분야(domain)에 따라 분류 성능이 크게 좌우되므로, 분야별로 추가 학습한 다양한 BERT가 공개되었다. 예를 들어 1,800백만건의 의학·생명학 논문으로 추가 학습한, 의학과 약학 분야의 BioBERT가 대표적이다. 금융, 경제 분야에서는 약 20만건의 미국 상장기업의 연차(10-K) 및 분기(10-Q) 보고서,⁴⁸⁾ 약 14만건의 기업실적보고회 원고(earnings call transcript), 약 49만건의 S&P 기업분석보고서 등에서 얻은 49억개의 단어로 학습한 FinBERT가 있다. FinBERT의 경우 사전학습만으로 구성된 BERT⁴⁹⁾에 비해 최대 16.1%의 성능 향상을 보여준다.⁵⁰⁾ 본 연구에서는 한글 위키피디아 자료로 사전 훈련된 KR-BERT를 설문의 내용으로 추가 학습⁵¹⁾하여 감성분석기를 구축하였다.

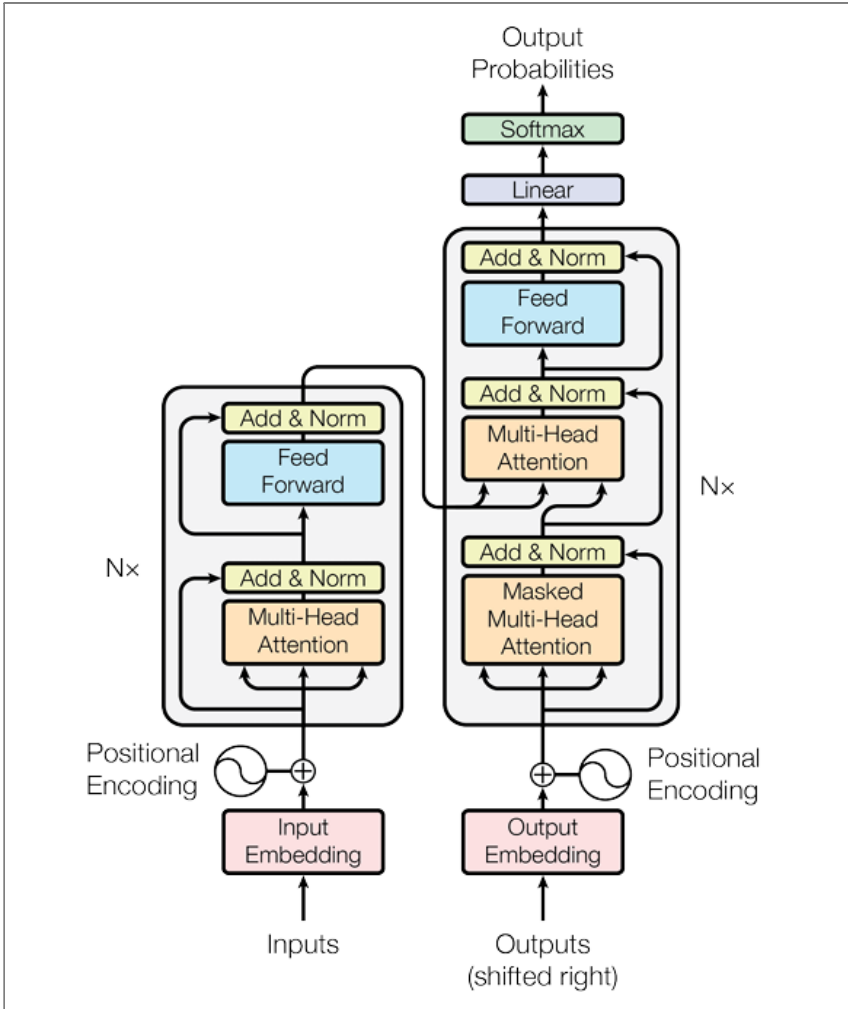
48) 미국 증권감독위원회(SEC)에 공개된 상장기업의 의무보고서로 기업의 영업 및 재무상태가 상세히 기술되어 있다. FinBERT에서는 영업활동(Item 1), 위험요소(Item 1A), 경영 및 분석(Item 7) 섹션을 활용하였다고 되어 있다.

49) 분야별 BERT와 구분하기 위해 generic-BERT라고도 한다.

50) BERT의 cased(대소문자를 구분하는 모형) 모형 기준이며 모형의 성능을 정확도(accuracy)로 측정한 경우.

51) 설문의 내용은 비교적 정형적인 문장과 단어로 구성되어 검열을 피해 변형된 단어들이 즐비한 인터넷 댓글과는 단어와 문장구조 측면에서 차이가 있다. 그런데도 검열을 피해 변형된 단어들은 주로 욕설에 해당하며 이들은 대부분 부정적인 감성을 띄게 되므로, 설문의 내용으로만 학습한 감성분석기를 사용한 경우도 본 연구의 분석 결과와 결론에는 영향이 없다.

[그림 V-1] 트랜스포머 모형 구조



자료: Vaswani et. al.(2017), p. 3

2) 학습자료의 전처리

학습자료는 3회에 걸친 개방형 설문조사이며 학습에 앞서 자료의 특성을 살펴볼 필요가 있다. 설문은 감성분류기의 학습을 염두에 두고 설계하였으므로, 긍정과 부정의 이유는 1~3개의 문장으로 답변하도록 하였다. 설문에

대해 피설문자가 대체로 성실하게 답변한 결과, 양질의 학습자료를 구성할 수 있어 전처리와 학습과정은 순조로웠다. 그러나 일부 불성실한 답변의 경우 감성분류기의 성능을 저하시키는 원인이 되므로 학습자료에서 제외하는 것이 분류의 편의(bias)를 줄이는 방법이다.⁵²⁾ <표 V-5>에는 설문조사 결과 암호화폐 과세에 대한 긍정의 의견을 밝힌 피설문자가 긍정의 근거로 제시한 답변 중 실질적인 근거가 될 수 없는 답변을 열거하였다. 이들은 대부분 ‘이유가 없다, 모른다’ 등 구체적인 이유가 명시되어 있지 않거나 암호화폐 과세 찬반 이유와 관련 없는 답변들이 대부분이다. <표 V-6>에서는 같은 이유로 학습자료에서 제외된 부정에 대한 근거들을 나열하였다. 부정 의견에서 제외된 답변은 긍정에 비해 수가 적지만 ‘탈세’, ‘가상’ 등 긍정 또는 부정의 의견을 뒷받침하지 못하는 단어만 제시한 경우 또는 ‘더 올려야 한다’와 같이 긍정의 의견을 제시한 경우로 학습자료에서 제외하였다. 이처럼 긍정과 부정의 감성을 분류하는 이유가 될 수 없는 답변은 학습의 성능을 크게 저하시키는 이상치(outlier)와 같은 역할을 한다. 따라서 이들을 사전에 제거함으로써 학습된 모형의 정확도(precision) 등을 향상시킬 수 있다.

설문자료를 활용하여 모형을 학습하는 데 다른 장애요인은 긍정과 부정의 비율이 큰 폭의 차이를 보이는 불균형(unbalanced) 자료라는 점이다. 학습자료의 긍정과 부정 비율이 큰 폭으로 차이를 보일 경우 비율이 높은 쪽으로만 학습해도 최소한 그 비율만큼의 정확도를 나타내므로, 학습의 결과가 코너해(corner solution)로 귀결될 가능성이 매우 높다. 3차에 걸친 설문자료에서 모두 70% 수준의 긍정 비율이 나타나므로, 모형은 모든 문장을 긍정으로만 분류해도 최소 70%의 정확도를 갖게 된다. 이 경우 부정으로 분류해야 할 문장을 긍정으로 분류함으로써 모두 ‘거짓 긍정’으로 분류되어 실질적으로 분류를 할 수 없는 모형이 학습된다. 따라서 자료를 균형 있는 자료로 보완해야 할 필요가 있는데, 이는 무작위 배치추출(random batch sampling) 방법⁵³⁾을 사용하였다. 무작위 배치추출의 경우 머신러닝에서 학습의 효율을

52) 이는 이상치(outlier)를 제거하는 winsorizing과 비슷하다. 이상치의 존재는 모형 추정 결과에 심각한 편의를 일으키기도 한다.

높이기 위해 학습자료를 보강하는 데 주로 사용되는 방법이다. 본 연구에서는 전체 답변의 30%에 불과한 부정 답변에서 중복이 허용되는 무작위 추출 방법으로 긍정과 부정의 비율이 50% : 50%가 되는 학습자료를 구성⁵⁴⁾하였다.

이상과 같이 마련된 텍스트 자료는 어간 추출(stemming), 표제어 추출(lemmatization), 불용어(stopwords) 제거 등의 전처리 과정을 거쳐 토큰화(tokenizing) 하고 품사(part-of-speech, POS) 꼬리표 달기(tagging)와 형태소 분석(morpheme analyzing)을 통해 분석을 위한 자료 준비를 마무리한다. 어간 추출에서 단어의 접두어, 접미어, 어미 등을 제거하고 의미를 내포한 어간만 추출하며, 표제어 추출에서는 변형된 동사, 형용사, 부사를 원형으로 복원한다. 불용어 제거 시에는 단어 자체만으로는 의미를 내포하지는 않으나, 상용어로 쓰이는 말(이, 그, 저, the 등)을 자료에서 제거한다. 품사 꼬리표⁵⁵⁾는 감성분석에서 주로 활용되는 명사(noun), 동사(verb), 형용사(adjective), 부사(adverb), 부정어(negation)를 추출하기 위해 꼭 필요한 작업이다. 이후에는 한글의 특성인 형태소 추출⁵⁶⁾을 거쳐야 한다. 이를 위한 기본 형태소 분석기는 python의 mecab library⁵⁷⁾를 사용하였다.

-
- 53) 통계학과 계량경제학에서 추정치의 통계적 검증을 위해 활용되는 bootstrapping 방법과 유사하다. 머신러닝에서도 무작위 배치추출을 통해 샘플의 크기를 늘리는 것만으로도 학습 결과가 개선되는 경우가 많다.
 - 54) 불균형자료의 단점을 보완하기 위해 학습가중치를 부여하는 방법 등도 고려할 수 있으나, 본 연구에서 사용된 자료에서는 편의가 없는(unbiased) 학습가중치를 부여할 기준을 찾지 못하였다. 따라서 무작위 추출로 표본의 수를 늘리는 방법으로 학습하였다.
 - 55) 추출되는 품사의 상세 내역은 〈부록 3〉을 참조.
 - 56) 한글의 특성상 형태소 추출에서는 실질형태소, 형식형태소, 자립형태소, 의존형태소로 구분되며, 이 중 실질형태소만 추출하는 과정에서 어간 추출, 표제어 추출이 동시에 이루어진다. 또한 명사, 동사, 형용사, 부사가 실질형태소에 해당하므로, 형태소 추출에서 품사 꼬리표까지 하나의 과정으로 이루어진다고 이해할 수 있다.
 - 57) 현재까지 개발된 한글 형태소 분석기에서 가장 개선된 성능을 보여준다.

〈표 V-5〉 ‘긍정’에서 제외된 답변

구분	제외한 답변	
긍정	<p>없음. 신선함이 가득함. 가상자산. 모르겠음. 굿굿, 창의적인 서비스입니다. 정날 다양하고 트렌디하다 현재는 그에대한불만이없음 이용이 쉽다 좋으니까 스토리 손쉽게양도가능 펼리할거 같다 돈의가치라생각 좋다고 생각한다 특별한 의견이 없습니다 편하다 글세요 잘모름 너무나도 없는 것 같다 명분 잘모르겠음 모름 새로운 화폐 신세대적이다 편리한점이 기억에 남음 취득세 난하지않으니까 흥미롭다 그냥그랬습니다 긍정적인 생각 좋은거 같아서 똑같은 돈이다. 긍정 신세계 글세요 소신적인 자발적 참여 좋아요 코인이 없는 것 같다. 특별히 없음 이유 없음 그만큼 유용해졌으니따 편하고 이대로 쪽 가격이저렴해서좋다 없습니다 좋습니다 딱히 없다.</p>	<p>너무좋아서여 가격이저렴해서좋다 돈을 많이 번다 그냥 처냥 가상자산을 안해서 별다른 이유는 없다. 잘 알지 못해서 좋아서 기술력 진보 투자 그냥 자산증가 과세열거방식 과세방안 없습니다 잘 모르겠다 투자해보았다 발전 가능성이 있다 다양한 소득 어어어어어 어어어어 웬지 바쁜일상에 편리할듯합니다 원래있었던데로 간편하다 안정적이라서 그냥 관심 없음스토리 그냥 ..? 특별한 생각없음 도있는사람들이하니까 매우 경제적이어서 글세요..... 긍정 잘 모르겠음 모르겠음 비싸다 효율성이 있다 너무 좋아요 처음들어봄 너무좋아서여 좋습니다 글세요 신 화폐 나쁘게생각해본적이없다 특별한 의견이 없습니다. 나는 안해서잘모르겠음 재산 증식을위한 세금 부여</p>

〈표 V-6〉 ‘부정’에서 제외된 답변

구분	제외한 답변	
부정	그냥 특별히 없음 잘모르겠음 굳이인느낌 신기루 잘 모른다좋습니다 모르겠음 투기 세금 재산이다 더 올려야 한다 특별히 없음 좋습니다	투기 자산 굳이? 잘모름 가상 특별히 없음 탈세 투기 조작 애매하다 잘몰라요..... 투기성 그냥 모르겠음

자료: 저자 작성

3) 감성분류기의 학습결과와 성능의 검정(validation)

머신러닝으로 학습한 모형은 모형의 복잡성과 비선형성으로 인하여 이론적으로 정립된 검정통계량으로 검정이 어렵다(Buckmann and Joseph, 2022). 따라서 학습된 모형을 평가하기 위해 정확성, 정밀도 등 예측 및 분류 성능을 평가하는 지표를 산정하여 비교한다. 우선 성능을 평가하기 위한 지표를 산정하기 위해 [그림 V-2]와 같이 모형의 분류 결과와 참값을 비교하는 혼동행렬(confusion matrix)을 구성하고 참과 거짓의 비율 등을 계산한다.

[그림 V-2] 혼동행렬(Confusion Matrix)

		Predicted	
		긍정	부정
Actual	긍정	참 긍정 (true positive)	거짓 부정 (false negative)
	부정	거짓 긍정 (false positive)	참 부정 (true negative)

자료: 저자 작성

이상의 혼동행렬부터 측정할 수 있는 평가지표는 다음과 같이 정확도, 정밀도, 재현율, F-1 score가 있다. 정확도(accuracy)는 분류의 정확도를 측정하는 척도로, 총 표본 수에서 참 긍정(true positive) 또는 참 부정(true negative)의 비율로 혼동행렬을 활용하여 이하와 같이 측정한다.

$$\text{Accuracy} = \frac{\text{true positive} + \text{true negative}}{\text{true positive} + \text{false positive} + \text{true negative} + \text{false negative}}$$

정밀도(precision)는 감성분류기가 긍정으로 분류한 표본 중 참 긍정을 긍정으로 옳게 분류한 비율이다. 재현율(recall)의 경우 감성분류기가 참 긍정 표본 중 긍정으로 옳게 분류한 비율을 나타낸다. F-1 score의 경우 정밀도와 재현율을 하나의 지표로 나타낸 점수이다.

$$\text{Precision} = \frac{\text{true positive}}{\text{true positive} + \text{false positive}}$$

$$\text{Recall} = \frac{\text{true positive}}{\text{true positive} + \text{false negative}}$$

$$\text{F1 score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

〈표 V-7〉 표본별 크기와 분류 성능 평가지표

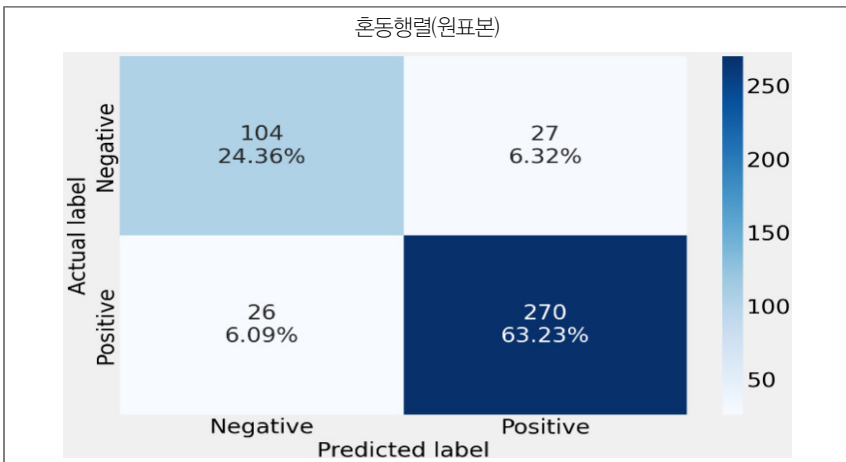
	원표본	수정표본	증강표본
답변 수			
긍정(개)	1,477	1,364	1,364
부정(개)	655	617	1,364
합(개)	2,132	1,981	2,728
분류성능 평가지표			
정확도	0.86	0.89	0.93
정밀도	0.87	0.89	0.93
재현율	0.86	0.89	0.93
F-1 score	0.86	0.89	0.93

주: 원표본은 개방형 설문조사의 답변으로 구성되어 있으며, 수정표본은 긍정 또는 부정으로 답변한 바와 일관되지 않은 서술 또는 의미 없는 서술을 제외한 표본이고, 증강표본의 경우 부정 답변의 무작위 중복 추출로 긍정과 부정 답변 수를 일치시킨 표본을 의미

자료: 저자 작성

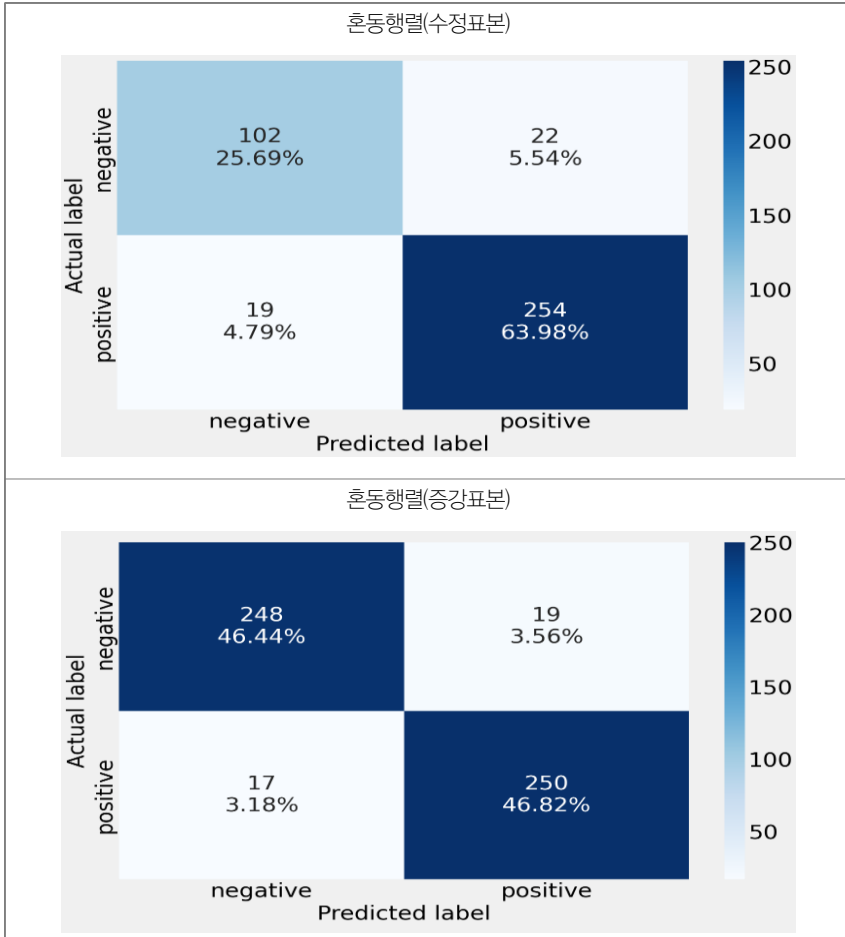
앞서 기술한 바와 같이 감성분류기를 학습하기 위한 표본은 무작위 중복 추출(random batch sampling)을 통해 긍정과 부정의 답변 수가 같아지도록 만든 표본을 사용하였다. 이를 증강표본이라고 할 때 설문 결과를 전처리 없이 사용하는 원표본과 모순되거나 의미가 없는 문장으로 모형의 성능을 저하시킬 수 있는 답변들을 제외한 수정표본과의 학습결과를 비교해 볼 수 있다. <표 V-7>은 각 표본의 수와 학습결과를 비교한 것이다. 원표본의 경우 긍정 1,477개, 부정 655개로 긍정:부정이 69.3:30.7의 비율이며, 수정표본은 68.9:31.1로 두 표본 모두 긍정 비율이 70%에 가깝다. 그런데도 학습 결과 산정되는 평가지표는 높은 수준인데, 원표본의 경우 정확도, 정밀도, 재현율, F-1 score가 각각 0.86, 0.87, 0.86, 0.86으로 나타났고, 수정표본도 네 종류의 평가지표가 모두 0.89로 나타났다. 수정표본에서 샘플링을 통해 긍정과 부정 비율을 각각 50%로 맞춘 증강표본은 네 지표가 0.93으로 나타나, 앞서 논의한 두 표본에 비해 성능이 크게 개선⁵⁸⁾되었음을 알 수 있다.

[그림 V-3] 표본 설정에 따른 혼동행렬



58) 일반적으로 텍스트 자료의 경우 측정오차로 인해 네 지표가 1의 값에 가까울수록 분류기의 성능 개선은 좀처럼 이루어지기 힘들다. 그런데도 증강표본으로 학습한 결과, 앞선 두 표본의 학습 결과에 비해 최소 0.4 이상의 성능 개선을 보이므로 성능이 대폭 개선되었다고 볼 수 있다.

[그림 V-3]의 계속



자료: 저자 작성

3. 감성사전을 활용한 인터넷 여론의 감성 측정

앞서 3회차에 걸친 가상자산 기타소득세 신설에 관한 개방형 설문조사 자료를 활용하여 트랜스포머 모형으로 구축한 감성분류기로 가상자산소득 과세 관련 뉴스의 댓글의 감성을 분석하고, 분석 결과를 개방형 설문조사 자료와 비교함으로써 뉴스의 댓글이 국민의 평균적인 조세 감정을 대표하는지

검증한다. 감성분류기로 뉴스 댓글의 감성을 분석하기에 앞서 감성분류기가 사전 분류된 댓글의 긍·부정 감성을 얼마나 잘 분류하는지를 점검한다. 만약 감성분류기가 댓글의 감성을 비교적 잘 측정하고 있다면, 오차 범위 내에서 댓글의 긍·부정 비율이 3회에 걸친 개방형 설문조사 자료의 결과와 근접하는지를 점검함으로써, 인터넷 여론이 국민의 조세 감정을 잘 대변하고 있는지 검증해 볼 수 있다.

가. 뉴스 댓글을 활용한 감성사전의 검증

수집한 댓글은 앞서 분석한 개방형 설문조사 자료와 같이 전처리를 통해 감성분석 자료로 준비하였다. 댓글의 감성을 분석하기에 앞서 설문자료로 학습한 감성분석기가 댓글의 감성을 잘 분석하고 있는지 검증할 필요가 있다. 학습자료와 검증자료의 성격이 크게 다를 경우 감성분석기의 정확도 등 평가지표도 하락할 수 있기 때문이다.

검증 방법은 수집한 댓글 중 123개를 직접 읽고 긍정·부정을 분류한 후 동일한 댓글을 구축된 감성분류기로 분류하였다. 분류 결과는 <표 V-8>의 혼동행렬과 같다. 우선 123개 댓글 중 부정 댓글 94개 중 90개(95.7%)를 부정으로 분류하였다. 반면 긍정 댓글 29개 중 14개(48.3%)만을 긍정으로 분류하였다. 설문자료로 학습한 감성분류기로 댓글을 분류할 경우 부정 댓글은 비교적 높은 정밀도로 분류하는 반면, 긍정 댓글의 경우 긍정으로 분류할 수 있는 댓글은 전체의 절반 수준이었다.

<표 V-8> 뉴스 댓글 분류 혼동행렬

(단위: 개)

Predicted \ Actual		긍정	부정
		긍정	14 (48.3)
부정	4 (4.3)	90 (95.7)	

주: () 안은 비중을 나타냄

자료: 저자 작성

감성분류기가 인터넷 댓글의 긍정 댓글의 절반을 부정으로 분류한 데는 <표 V-9>와 같이 세 가지 이유로 분석해 볼 수 있다. 첫째, 욕설 또는 비속어와 같이 강한 부정적 감성을 지닌 단어가 쓰이고 있는 경우이다. 해당 단어에 상대적으로 많은 가중치가 부여될 때 문장 전체가 부정 감성으로 분류되는 경우가 발생하는 것이다. 둘째, 부정에 대한 부정으로 긍정을 표현한 댓글의 경우이다. 부정의 부정 문제는 감성분석에서 흔히 거론되는 이슈로, 두 부정 중 하나만 부정으로 인식될 때 긍정이 부정으로 인식되는 오류가 발생한다. 세 번째는 문장에 긍정·부정을 나타내는 단어는 없지만 문맥으로만 긍정·부정을 분류할 수 있는 경우이다. 흔히 말하듯 ‘강한 부정은 긍정과 같은 경우⁵⁹⁾’로 감성분류기의 경우 부분에 대한 학습이 필요하나, 문맥만으로 긍정·부정을 판별할 정도의 학습이 진행될 경우 과적합(over-fitting)의 문제가 발생할 우려가 커진다. 따라서 부정으로 분류된 긍정 댓글을 수정하기 위한 추가 학습과정이나 미세조정 등을 거치지 않는 것이 모형의 일치성(consistency)을 높이는 방법이다. 인터넷 댓글의 감성분류 시 이를 고려하여, 감성분류기로 분류된 긍정 댓글 비율의 두 배 비율까지도 긍정적인 댓글이 있을 수 있다는 점을 고려하여 분석할 것이다.

<표 V-9> 부정으로 분류된 긍정 댓글

구분	댓글	분류 확률		분류
		긍정	부정	
비속어 등 부정적 감성을 지닌 단어가 포함된 경우				
1	소득이 있으면 세금을 내야지. 씨밤 여기가 북한이냐??	0.23	0.77	부정
2	세금70프로 때려	0.04	0.96	부정
3	코인과세 당근해야지 조빵이치며 돈버는 서민들은 세금 따박따박 내는데 클릭질 몇번해서 세금도 안내고 돈벌게?	0.04	0.96	부정
4	그니까 암호화폐 하지 말고 주식하라고... 암호화폐 소득에 세금혜택 주는 건 미친 짓이지	0.03	0.97	부정

59) 예를 들어 단풍이 든 산을 보며 “이야 경치 죽이는데~!”라고 하면 사람은 이를 매우 긍정적으로 인식하지만, 감성분류기의 경우 ‘죽이다’라는 말에 많은 가중치가 더해지며 부정으로 분류되기 쉽다.

〈표 V-9〉의 계속

구분	댓글	분류 확률		분류
		긍정	부정	
5	코인은 뭐하는데 세금혜택을 준다는 것이냐? 악덕 코인들 당장 없애라. ... 노동자와 서민들이 누들 봉이나? 우리들한테 걷은 세금으로 별 그지 같은 사기에 세금혜택을 부여해 달라고 하고 있어. 가면 있으니 누굴 봉으로 아냐. 한국의 노동자들이여, 한국의 월급쟁이들이여, 한국의 서민들이여 단결하여 저 간악한 코인 투자자들, 코인 발행자들에게 죽창을 날리자.	0.04	0.96	부정
6	비과세를 왜함 과세는 평등해야하지 표연으려고 쌍 난리 피우지말고 싹 다 과세해라	0.11	0.89	부정
부정에 대한 부정으로 긍정을 표현한 경우				
7	비트코인 뭐라고 세금을 안 내는거야	0.03	0.97	부정
8	대다수의 국민은 민주당 응원하고, 소수를 위한 과세연기 반대한다	0.08	0.92	부정
9	요계 맞지. 주식이야 기업에 투자 개념이지만 코인은 사실상 투기인데 이런 자산에 비과세 혜택을 주는건 이상한거지	0.04	0.96	부정
10	만약 코인과세를 하지 않으면...코인은 제도권으로 들어오지 않고, 피해자들이 생긴다. 그리고, 세금을 월급쟁이들이 다낸다	0.04	0.96	부정
11	대다수의 국민은 민주당 응원하고, 소수를 위한 과세연기 반대한다	0.08	0.92	부정
단어 조합이 아닌 행간의 의미로 긍정/부정을 판단할 수 있는 경우				
12	민주당 말 함 들어주고 과세가세요 솔직히 코인도 엄연한 투자잖아요 손해를 감수하고 이득본거에 대한 세금내라고 하는건데...	0.04	0.96	부정
13	코인은 무법지대의 주식임 빨리 코인 자체를 도박 및 불법화 하던가, 아니면 탈중앙화 버리고 세금 메겨서 법제화 하던가.	0.06	0.94	부정
단순 오류				
14	세금걸자	0.18	0.82	부정

자료: 감성분석 결과를 바탕으로 저자 작성

나. 댓글의 감성분류와 비교 분석

앞서 기술한 바와 같이 3회에 걸쳐 독립적으로 시행된 개방형 설문조사에서 가상화폐 과세에 대한 긍정 답변 비율은 69.3%에 이르며, 이 비율이 일관되게 나타난다. 그렇다면 만약 뉴스에 대한 댓글들이 국민들의 평균적인 조세 인식을 잘 반영한다면, 뉴스에 대한 댓글의 감성을 측정할 때도 유사한 비율이 나타날 것을 기대할 수 있다. 따라서 이 검증 결과를 고려하여 전체 댓글을 긍정과 부정으로 분류하였을 때 개방형 설문조사의 결과와 유

사한지 아닌지를 비교하면, 뉴스에 대한 댓글이 국민의 평균적인 여론을 잘 반영하는지를 확인할 수 있다.

댓글은 시간에 따른 과세에 대한 인식 변화를 고려하여 연도별로 구분하여 분류하였으며, 결과는 <표 V-10>과 같다. 2018년부터 2022년 6월까지 긍정으로 분류된 댓글의 비율은 각각 0.09, 0.08, 0.08, 0.08, 0.07로 평균 8% 댓글만 긍정으로 분류되었다. 본 연구에서 구축된 감성분류기가 약 50% 확률로 긍정 댓글을 긍정으로 분류한다는 앞의 검증 결과를 반영한다고 하더라도 전체 댓글의 16%만이 긍정 댓글이라고 평가할 수 있다. 반면 부정 댓글은 5년간 전체 댓글 중 90% 이상이었다. 검증 결과를 고려하여 조정할 때도 80% 이상의 비율이다.

댓글의 긍정과 부정의 비율은 댓글에 대해 독자들이 표시하는 ‘좋아요’와 ‘싫어요’를 가중치로 하여 측정하였을 때 더욱 큰 폭의 차이를 보여준다. 가중치를 고려한 긍정·부정 댓글의 수는 N_p , N_n , N 을 각각 긍정 댓글 수, 부정 댓글 수, 총 댓글 수라고 할 때 다음과 같이 구할 수 있다.

$$\text{Weighted Positive} = N_p + \sum_{i=1}^N I_p(\text{Comment}_i) \times (\text{Like}_i - \text{Dislike}_i)$$

$$\text{Weighted Negative} = N_n + \sum_{i=1}^N I_n(\text{Comment}_i) \times (\text{Like}_i - \text{Dislike}_i)$$

여기에서 $I_p(\text{Comment}) = 1$ if $\text{Comment} \in \{\text{Positive Comments}\}$,

$I_n(\text{Comment}) = 1$ if $\text{Comment} \in \{\text{Negative Comments}\}$

가중치를 고려한 긍정과 부정의 댓글 비중은 각각 4%, 96%가 되어 앞선 비교 결과보다 더욱 큰 폭의 차이를 보인다. 이는 긍정적으로 쓴 댓글보다 부정적으로 쓴 댓글이 훨씬 많은 ‘좋아요’를 받았다는 것인데, 이는 인터넷 여론을 작성하는 이들 대부분이 해당 조세정책에 부정적인 의견을 가진 사람이었을 것임을 추측할 수 있게 해준다. 이는 인터넷 여론에 기반한 정책 피드백 분석이 국민의 평균적 조세 인식을 왜곡하여 반영할 수 있다는 사실을 입증한다.⁶⁰⁾

〈표 V-10〉 연도별 댓글의 감성

(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: 저자 작성

이러한 분석 결과는 유튜브 플랫폼에 기록된 댓글에서도 동일하게 나타난다. 제IV장 제2절에서 분석한 유튜브 댓글 총 4,579건을 감성분류한 결과, 긍정 607건(13.3%), 부정 3,972건(86.7%)으로 분류되었다. 댓글 감성분류의 검증 결과, 참 긍정을 긍정으로 분류할 확률이 50% 수준이라는 점을 고려하더라도 긍정 댓글의 비율은 27% 이하이다. 이는 네이버 플랫폼에 비해서는 높은 수치이나 설문 결과 나타난 수치의 절반에도 미치지 못하는 수준임을 알 수 있다.

4. 설문조사 응답과 네이버 댓글 감성별 텍스트 심층 분석

본 소절에서는 네이버의 가상자산 과세 관련 2022년 기사 댓글 중 긍정적 문장과 부정적 문장을 설문의 긍정 응답 및 부정 응답과 비교하여 분석함

60) 『연합뉴스』, 「선거 앞두고 1년 밀린 가상자산 과세...세금은 2024년부터 낸다」, 2021. 11. 30., <https://www.yna.co.kr/view/AKR20211130122400002>, 검색일자: 2022. 10. 5.

로써 설문 응답과 댓글의 차이를 보다 심도 있게 분석한다. 지나치게 짧은 문장 제거 등을 거친 결과, 설문 응답 분석 표본은 1,231개의 긍정 문장, 390개의 부정 문장으로 구성되었으며, 댓글 분석 표본은 130개의 긍정 문장, 982개의 부정 문장으로 이루어져 있다.

〈표 V-11〉은 긍정·부정 감성별로 댓글과 설문의 사용 어휘를 비교하기 위하여 각 데이터에서 크기가 1인 고빈도 단어 집합을 나열한 결과를 보여 준다. 긍정적 댓글의 경우 ‘공매도’에 대한 비판, ‘윤석열’ 대통령(후보)에 대한 비판 등이 과세 긍정으로 이어졌지만, 긍정 설문 응답의 경우 ‘불로소득’에 대한 과세, 과세 ‘형평성’에 대한 고려, ‘탈세’에 대한 우려 등이 긍정적인 의견을 주도하고 있다. 부정적 댓글의 경우 ‘민주당’에 대한 비판의 빈도가 높았으며, 부정 설문 응답의 경우 투자자 ‘보호’에 우선할 것과 정당한 수익을 ‘인정’할 것, ‘과도한 ‘세율’에 대한 경계 등이 반영되어 있다. 앞 장에서 살펴본 바와 유사하게 긍정 부정 여부와 상관없이 댓글의 경우 정치적인 견해가 많이 반영되어 있음을 확인할 수 있다. 다만 감성으로 구분한 결과, 특정 단어 집합이 어떠한 이유로 긍정 혹은 부정적인 의견을 이루는지가 더욱 선명하게 보이는 것으로 나타났다.

〈표 V-11〉 감성별 2022년 네이버 댓글과 설문 응답의 크기 1인 고빈도 단어 집합

긍정 문장				부정 문장			
2022년 네이버 댓글		설문 응답		2022년 네이버 댓글		설문 응답	
단어 집합	지지도	단어 집합	지지도	단어 집합	지지도	단어 집합	지지도
{‘주식’}	0.131	{‘소득’}	0.468	{‘세금’}	0.145	{‘과세’}	0.376
{‘폐지’}	0.115	{‘과세’}	0.367	{‘민주당’}	0.098	{‘가상’}	0.363
{‘공매도’}	0.108	{‘가상’}	0.345	{‘코인’}	0.095	{‘자산’}	0.329
{‘윤석열’}	0.100	{‘자산’}	0.344	{‘주식’}	0.084	{‘세금’}	0.267
{‘대통령’}	0.069	{‘세금’}	0.273	{‘공약’}	0.079	{‘소득’}	0.182
{‘공약’}	0.069	{‘발생’}	0.105	{‘과세’}	0.074	{‘화폐’}	0.147
{‘세금’}	0.062	{‘부과’}	0.091	{‘국민’}	0.071	{‘부과’}	0.116
{‘자산’}	0.062	{‘투자’}	0.087	{‘윤석열’}	0.068	{‘정부’}	0.104
{‘가상’}	0.062	{‘화폐’}	0.085	{‘사람’}	0.067	{‘투자’}	0.104
{‘양도세’}	0.062	{‘거래’}	0.070	{‘폐지’}	0.066	{‘보호’}	0.088

〈표 V-11〉의 계속

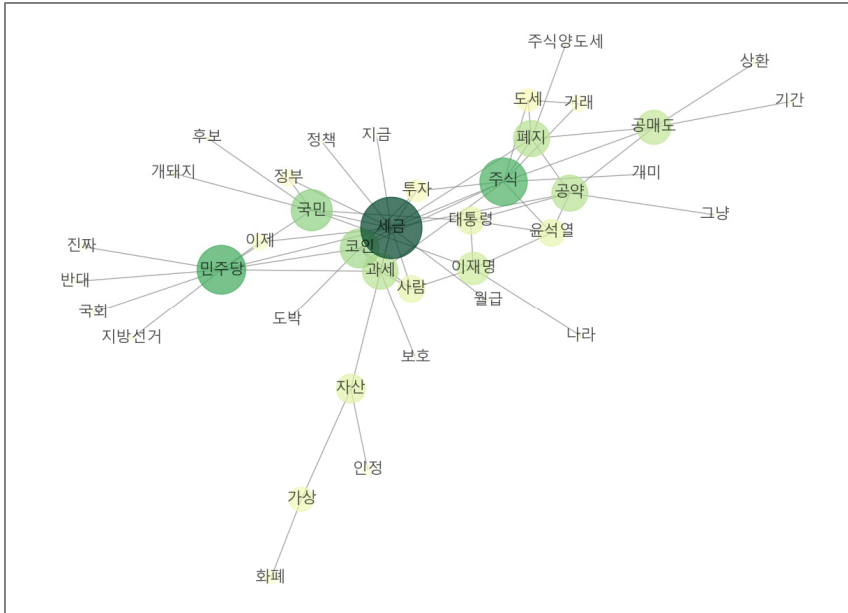
긍정 문장				부정 문장			
2022년 네이버 댓글		설문 응답		2022년 네이버 댓글		설문 응답	
단어 집합	지지도	단어 집합	지지도	단어 집합	지지도	단어 집합	지지도
{‘정책’}	0.054	{‘주식’}	0.062	{‘이재명’}	0.065	{‘세울’}	0.073
{‘국민’}	0.046	{‘모든’}	0.058	{‘대통령’}	0.057	{‘주식’}	0.071
{‘코인’}	0.038	{‘투기’}	0.056	{‘가상’}	0.054	{‘부정’}	0.065
{‘인정’}	0.038	{‘불로소득’}	0.052	{‘양도세’}	0.050	{‘손실’}	0.065
{‘위해’}	0.038	{‘수익’}	0.050	{‘정부’}	0.047	{‘제도’}	0.057
{‘과세’}	0.038	{‘형평성’}	0.039	{‘나라’}	0.043	{‘인정’}	0.055
{‘투자’}	0.031	{‘내야’}	0.039	{‘공매도’}	0.043	{‘거래’}	0.055
{‘부자’}	0.031	{‘적용’}	0.037	{‘자산’}	0.041	{‘개인’}	0.053
{‘활성화’}	0.031	{‘세울’}	0.037	{‘화폐’}	0.040	{‘적용’}	0.045
{‘해외’}	0.031	{‘사람’}	0.035	{‘그냥’}	0.034	{‘수익’}	0.039
{‘지지’}	0.031	{‘방지’}	0.033	{‘후보’}	0.033	{‘국가’}	0.037
{‘거래’}	0.031	{‘공정’}	0.033	{‘투자’}	0.033	{‘정책’}	0.037
{‘소득’}	0.031	{‘탈세’}	0.032	{‘진짜’}	0.033	{‘과도’}	0.035

자료: 분석 결과를 바탕으로 저자 작성

다음으로 의미망 분석을 진행하였다. [그림 V-4], [그림 V-5]는 각각 설문 응답과 네이버 댓글의 긍정 문장 의미망을, [그림 V-6], [그림 V-7]은 각각 설문 응답과 네이버 댓글의 부정 문장 의미망을 시각화하고 있다. 제IV장 제2절의 분석에서 본 바와 같이 설문 응답의 경우 주요 어휘들이 하나의 클러스터에 밀집되어 있다. 긍정 응답의 경우 세금 부과와 형평성, 투기로 인한 이익에 대한 과세, 불로소득에 대한 과세 관련 어휘 등이 발견되며, 부정 응답의 경우 투자자 보호와 과도한 세율 적용 관련 어휘가 발견된다. 댓글 어휘의 의미망은 설문 응답 어휘에 비해 산개되어 있으며 다수의 클러스터가 발견된다. 긍정 댓글의 경우 ‘형평성’ 고려 등 과세의 필요성 관련 클러스터와 ‘공매도’ 비판 관련 클러스터가 정치 관련 클러스터와 함께 발견된다. 부정 댓글의 경우 ‘주식양도세’ 폐지 주장 관련 클러스터와 정치 관련 클러스터 등이 발견된다.

댓글의 경우 긍정 부정 댓글 의미망 모두에서 정치 관련 클러스터가 뚜렷

[그림 V-7] 가상자산 과세 관련 2022년 네이버 기사 부정 댓글의 의미망 시각화



자료: 분석 결과를 바탕으로 저자 작성

Ⅵ. 결론 요약 및 시사점

최근 발달하고 있는 텍스트 마이닝 기법은 뉴스나 댓글과 같은 텍스트 정보를 수량화된 배열로 나타내 분석하는 이론적 틀을 제시하여 텍스트를 정량적으로 분석할 수 있는 기반을 제공하고 있다. 이러한 텍스트 분석 방법론을 이용하여 황의찬·우석진(2022)은 조세정책에 대한 인터넷 미디어 여론의 부정적인 감성이 강해질수록 정책이 수정되거나 철회될 가능성이 크다는 것을 보여주었다. 만일 인터넷 미디어 여론이 조세정책에 대한 국민들의 선호를 잘 반영한다면 이러한 현상은 바람직할 수 있지만, 만일 그렇지 못한 경우 인터넷 여론을 의식하여 정책을 수정하거나 철회하는 것은 사회적 합의를 반복하는 것으로 공정하지 못하며 세제 안정성을 해칠 수도 있다.

본 연구에서는 인터넷 미디어 여론이 조세정책에 대한 국민적 선호를 잘 반영하고 있는지를 텍스트 마이닝 방법론을 통해 정성적·정량적으로 분석해 보았다. 이를 위해서 먼저 개방형 설문을 통해 가상자산 기타소득세에 대한 긍·부정 여부와 그렇게 생각하는 이유에 대해 설문하여 실제 국민들의 선호와 근접한 텍스트를 수집하였다. 그리고 이렇게 얻은 텍스트를 인터넷 미디어 여론(네이버 뉴스 기사 및 댓글, 유튜브 댓글)과 정성적·정량적으로 비교해 보았다.

먼저 제Ⅵ장에서 연관규칙학습을 통해 정성적으로 분석된 결과를 보면, 설문 응답의 경우는 고빈도 단어 집합의 지지도가 높아 사용되는 어휘가 정제되어 있고 조세정책과 밀접한 연관성이 있는 단어가 많이 사용되고 있었다. 반면 인터넷 미디어 여론의 경우 어휘의 지지도가 낮아 사용 어휘가 산발적으로 분포되어 있음을 알 수 있었으며, 정치 관련 어휘들이 많아 조세정책을 중심으로 형성된 여론이라고 보기 어려운 점이 존재하였다.

이후 제Ⅶ장에서 토픽 모델링과 감성분석을 통해 설문에서 얻어진 텍스트

와 인터넷 미디어 여론을 비교하였다. 우선 LDA 토픽 모델링을 하여 텍스트들을 구성하고 있는 주제들을 분류해 보았다. 토픽 모델링을 통해 분류해 본 결과, 인터넷 미디어 여론들은 분석대상이 되는 가상자산 기타소득세에 관한 주제 이외에도 여러 주제들을 포함하고 있는 것을 확인할 수 있었다. 예를 들어 2018년의 댓글 중 44%는 가상자산 기타소득세가 아닌 부동산에 대한 주제를 형성하고 있었다. 2019년 댓글도 34%는 전반적으로 증세 기조를 보이는 정부에 대한 불만을 주제로 하고 있었다. 2022년의 경우도 30%의 댓글은 금융투자소득세와 공매도에 관한 내용을 주제로 하고 있었으며, 27%의 경우에는 대선 후보들에 관한 내용을 주제로 하고 있었다.

반면 설문을 통해서 얻은 텍스트를 토픽 모델링을 통해 분석하였을 때는 전반적으로 조세정책과 밀접한 관련이 있는 주제를 형성하고 있는 것을 확인할 수 있었다. 긍정으로 응답한 텍스트들은 불로소득에 대한 과세(14%), 수익에 대한 과세(20%) 등 응답자들이 생각하고 있는 조세정책에 대한 원칙들을 주제로 하는 경우가 많았으며, 부정적으로 응답한 텍스트들의 경우 투자자 보호(12%)와 과도한 세율(13%), 위험 감수에 따른 정당한 수익(20%) 등을 주제로 형성하고 있었다.

감성분석을 위해서는 트랜스포머 모형을 이용하였으며 설문 텍스트를 학습 자료로 사용하였다. 학습된 감성분류기는 정확도, 정밀도, 재현율, F-1 score 에서 0.86~0.93 수준의 높은 성능을 나타내었다. 설문에서는 가상자산 기타소득세에 대한 긍정 비율이 69.3% 정도를 보였기 때문에, 인터넷 여론이 국민들의 선호를 잘 반영한다면 유사한 긍정 비율이 나와야 할 것으로 기대할 수 있다. 하지만 네이버 뉴스 댓글의 경우 약 8%만이 긍정으로 분류되었으며, 90% 이상이 부정으로 분류되었다. 분류기가 댓글에 대해서 잘못 분류할 오류의 가능성을 고려하더라도 긍정 분류는 16% 정도였으며, 부정 분류는 80% 이상으로 나타나 인터넷 미디어 여론은 조세정책에 대한 부정적인 여론을 매우 편향적으로 나타내고 있음을 보여주고 있다. 유튜브의 댓글도 긍정 비율이 더 높게 나타나기는 하였으나, 전반적으로는 설문 응답에 비해 부정적인 편향성이 있는 것을 확인할 수 있었다.

설문과 인터넷 여론이 크게 다를 수 있는 이유에 대해서 추가로 검토해 보았다. 우선 댓글의 긍·부정의 분류를 독자들이 댓글에 대해서 표기하는 ‘좋아요-싫어요’ 수로 가중치를 부여하는 작업을 수행하면, 긍정 분류가 4%가 되고 부정 분류가 96%가 되어 가중치가 없는 경우에 비해서 그 차이가 더욱 심화하였다. 이는 긍정적으로 쓴 댓글보다 부정적으로 쓴 댓글이 훨씬 많은 ‘좋아요’를 받았다는 것인데, 이는 인터넷 여론을 작성하는 이들 대부분이 조세정책에 부정적인 의견을 가진 사람이었을 것임을 추측할 수 있게 한다.

연관규칙학습과 의미망 분석을 통해서 설문의 긍·부정 응답과 인터넷 미디어 여론에서 긍·부정으로 분류되는 텍스트를 비교할 때도 두 텍스트의 특징은 크게 다른 것으로 나타났다. 설문의 텍스트들은 LDA 분석과 같이 긍정 분류에서는 이익에 대한 과세, 불로소득에 대한 과세와 관련된 어휘가 주로 발견되었으며, 부정 분류에서는 투자자 보호와 같은 어휘들의 출현 빈도가 높았다. 반면 댓글에는 긍정 분류의 경우 형평성과 공매도 비판이 발견되었으며, 부정 댓글에는 주식양도세 폐지 등이 발견되어 주제들이 산발적인 것을 확인할 수 있었으며, 조세정책과는 무관한 정치 관련 어휘들이 많이 발견되었다.

이상의 결과를 요약하자면 특정 조세정책에 대한 인터넷 미디어 여론은 텍스트의 주제가 매우 산발적이어서 조세정책에 대한 선호를 잘 반영하지 못할 것으로 판단할 수 있다. 또한 감성분석 결과는 인터넷 미디어 여론이 조세정책에 대한 부정적인 의견이 매우 편향적으로 표출하여 조세정책에 대한 국민들의 선호를 왜곡하여 반영할 가능성이 있음을 정량적으로 보여주고 있다. 이는 인터넷 여론을 의식하여 정책을 수정하거나 철회하는 것은 사회적 합의를 번복하는 것으로 공정하지 못하며 세계 안정성을 해칠 수도 있다는 주장을 뒷받침하는 결과이다.

본 연구는 특정 조세정책을 중심으로 수행된 기초연구로서 연구 결과를 일반화하기에는 추가적인 분석이 요구된다는 한계점이 있다. 하지만 가상자산 기타소득세 신설의 경우, 2020년 정부발의 세법개정안에서 시행 시기를

‘2021년 10월 1일’에서 ‘2022년 1월 1일’로 변경하여 통과시켰으며, 2021년 의원발의 세법개정안으로 2023년 1월 시행하는 것으로 유예되었고, 2022년 정부발의 세제개편안에서는 2025년으로 유예하는 것을 논의하고 있는 상황으로, 본 연구의 결과는 밀접한 관련이 있을 것으로 생각한다. 직접적인 시사점으로는 인터넷 미디어 여론은 해당 조세정책에 대한 국민적인 선호를 잘 반영하지 못하고 편향되어 있으므로, 이를 정책 수립과정에 여과 없이 반영하는 것에 대해서 경계해야 할 것이라는 메시지를 제공한다.

좀 더 장기적이고 일반적인 시사점으로는 조세정책 입안 과정에서의 충분한 사회적 합의와 세제 안정성과 예측성을 높일 수 있는 방안이 필요함을 제시한다. 이를 위해서는 경제학, 정치학 등 다양한 측면에서의 후속 연구가 필수적일 것이다. 구체적으로 국내 미디어 여론의 편향성에 관한 체계적인 연구, 정치적 양극화와 여론 양극화의 측정 및 그 기제에 관한 연구, 조세정책에 관한 국민들의 선호를 도출하고 통합할 수 있는 체제에 관한 연구 등 여러 분야에서 본 연구보다 더 넓은 범위에서 발전된 방법론으로 기초연구, 정책연구, 학술적 연구가 지속해 수행되기를 기대한다.

참고문헌

〈국내 문헌〉

- 국회예산정책처, 『2013년 세법개정안 분석』, 2013.
- _____, 『2014년 세법개정안 분석』, 2014.
- _____, 『2015년 세법개정안 분석』, 2015.
- _____, 『2016년 세법개정안 분석』, 2016.
- _____, 『트렌드 세법 -과거 20년간 세법개정안의 궤적을 담다』, 2017a.
- _____, 『2017년 세법개정안 분석』, 2017b.
- _____, 『2018 시행 개정세법의 주요 내용 및 심사 쟁점』, 2017c.
- _____, 『2019 시행 개정세법의 주요 내용 및 심사 쟁점』, 2018.
- _____, 『2020 시행 개정세법의 주요 내용 및 심사 쟁점』, 2019.
- _____, 『2020년 개정세법의 심사경과와 주요 내용』, 2020.
- _____, 『2021년 개정세법의 심사경과와 주요 내용』, 2021.
- 『경향신문』, 「대선 앞두고…세금 깎아줄게, 표 좀 나오?」, 2021. 11. 3.,
https://www.khan.co.kr/print.html?art_id=202111031508001, 검색일자:
2022. 10. 5.
- 기획재정부, 『2022년 세제개편안』, 2022.
- 기획재정부 보도자료, 「5,500만원 이하자 세법개정에 따른 세부담 증가 해소
- '15년 연말정산 보완대책 -」, 2015. 4. 6.
- _____, 「주택시장 안정대책」, 2018. 9. 13.
- _____, 「주택시장 안정화 방안」, 2019. 12. 16.
- _____, 「다주택자 양도소득세 중과 한시 배제 등 「소득세법 시행령」 개정」,
2022. 5. 9.
- 기획재정부 보도참고자료, 「신용카드 소득공제 관련 기재부 입장」, 2019. 3. 11.

- 길호현, 「텍스트마이닝을 위한 한국어 불용어 목록 연구」, 『우리말글』, 제78호, 2018, pp. 1~25.
- 김기성·김종순·홍근석, 「지방세 납세순응의 영향요인에 관한 연구: 부산광역시 납세자들의 인식을 중심으로」, 『지방행정연구』, 제28권 제4호, 2014, pp. 391~416.
- 김수현·이영준·신진영·박기영, 「경제 분석을 위한 텍스트 마이닝」, 『BOK 경제연구』, 제2019-18호, 한국은행, 2019.
- _____, 「거시경제 분석을 위한 텍스트 마이닝」, 『한국경제의 분석』, 제26권 제1호, 2020, pp. 1~70.
- 김재진, 「연말정산 대란과 보완대책, 그리고 남은 과제들」, 『재정포럼』, 6월 호(통권 241호), 한국조세재정연구원, 2016, pp. 6~33.
- _____, 「탈세, 세무조사가 유일한 해법인가?: Beyond the Allingham and Sandmo Model」, 『재정포럼』, 통권 제249호, 한국조세재정연구원, 2017, pp. 6~15.
- 김진주·윤종빈, 「텍스트마이닝을 통한 온라인 플랫폼에서의 이념 갈등 표출에 관한 연구: 제21대 총선 유튜브 개표방송의 라이브 스트리밍을 중심으로」, 『정치·정보연구』, 제24권 제1호, 2021, pp. 101~135.
- 박명호, 『우리나라 국민들의 납세의식 조사』, 한국조세재정연구원, 2015.
- _____, 「납세에 관한 일반 국민들의 인식 변화 분석」, 『재정포럼』, 통권 제246호, 한국조세재정연구원, 2016, pp. 26~44.
- 박명호·심태섭, 「우리나라 납세자의 납세순응태도에 영향을 미치는 요인」, 『세무와 회계저널』, 제9권 제4호, 2008, pp. 55~84.
- 박지용·김재진·구재이, 『세금, 알아야 바꾼다』, 2018.
- 『연합뉴스』, 「선거 앞두고 1년 밀린 가상자산 과세…세금은 2024년부터 낸다」, 2021. 11. 30., <https://www.yna.co.kr/view/AKR20211130122400002>, 검색일자: 2022. 10. 5.
- 『이데일리』, 「[집 Talk] 억울한 1주택자…“우리도 투기꾼이냐”」, 2020. 1. 24., <https://n.news.naver.com/mnews/article/018/0004561670?sid=101>, 검색일자: 2022. 10. 2.

- 이새미·유승의·안순재, 「텍스트 마이닝을 활용한 매스 미디어와 소셜 미디어 의제 분석: '마스크 5 부제'를 중심으로」, 『한국콘텐츠학회 논문지』, 제20권 제6호, 2020, pp. 460~469.
- 이환용·고창수·배진수, 『머신러닝을 활용한 재정사업평가: 정책금융 사례를 중심으로』, 한국조세재정연구원, 2021.
- 임원혁·이창근·최동욱·정세은, 『한국의 여론양극화 양상과 기제에 관한 연구』, 한국개발연구원, 2019.
- 홍성훈·노희천·정희선·김용수, 『납세의식 제고를 위한 조세행정 개선방향 연구』, 한국조세재정연구원, 2017.
- 황의찬, 「납세자의 정서가 조세정책에 미치는 영향」, 명지대학교 석사학위 논문, 2021.
- 황의찬·우석진, 「납세자의 정서가 정부의 조세정책 결정에 미치는 영향」, 『법경제학연구』, 제19권 제1호, 2022, pp. 229~263.

〈외국 문헌〉

- Agrawal, R. and Srikant, R., “Fast algorithms for mining association rules,” In Proceedings of the 20th international conference on very large data bases, *VLDB*, Vol. 1215, 1994, pp. 487~499.
- Agrawal, R., Imieliński, T., and Swami, A., “Mining association rules between sets of items in large databases,” In Proceedings of the 1993 ACM SIGMOD international conference on Management of data, 1993, pp. 207~216.
- Ali, M. M., Cecil, H. W., and Knoblett, J. A., “The effects of tax rates and enforcement policies on taxpayer compliance: A study of self-employed taxpayers,” *Atlantic Economic Journal*, 29, 2001, pp. 186~202.
- Allingham, M. G., and Sandmo, A., “Income tax evasion: A theoretical analysis,” *Journal of Public Economics*, 1(3-4), 1972, pp. 323~338.
- Alm, J., “Tax compliance and administration,” Working Paper, University

- of Colorado at Boulder, 1998.
- Alm, J., Cherry, T., Jones, M., and McKee, M., "Taxpayer information assistance services and tax compliance behavior," *Journal of Economic Psychology*, 31(4), 2010, pp. 577~586.
- Alm, J., Jackson, B. R., and McKee, M., "Fiscal exchange, collective decision institutions, and tax compliance," *Journal of Economic Behavior & Organization*, 22(3), 1993, pp. 285~303.
- Alm, J., McClelland, G. H., and Schulze, W. D., "Why do people pay taxes?," *Journal of public Economics*, 48(1), 1992, pp. 21~38.
- Alm, J., McClelland, G. H., and Schulze, W. D., "Changing the social norm of tax compliance by voting," *Kyklos*, 52(2), 1999, pp. 141~171.
- Alm, J., Sanchez, I., and De Juan, A., "Economic and noneconomic factors in tax compliance," *KYKLOS-BERNE*, 48, 1995, pp. 1~22
- Baker, S. R., Bloom, N., and Davis, S. J., "Measuring economic policy uncertainty," *The Quarterly Journal of Economics*, 131(4), 2016, pp. 1593~1636.
- Baldry, J. C., "Income tax evasion and the tax schedule: Some experimental results," *Public Finance= Finances Publiques*, 42(3), 1987, pp. 357~383.
- Becker, W., Büchner, H. J., and Sleeking, S., "The impact of public transfer expenditures on tax evasion: An experimental approach," *Journal of Public Economics*, 34(2), 1987, pp. 243~252.
- Beron, K. J., Tauchen, H. V., and Witte, A. D., "The effect of audits and socioeconomic variables on compliance," *In Why People Pay Taxes*, edited by J. Slemrod, 1992, pp. 67~89.
- Bohnet, I. and Frey, B. S., "Direct-democratic rules: The role of discussion," *Kyklos*, 47(3), 1994, pp. 341~354.
- Bosco, L. and Mittone, L., "Tax evasion and moral constraints: some experimental evidence," *Kyklos*, 50(3), 1997, pp. 297~324.

- Bott, K., Cappelen, A. W., Sørensen, E. Ø., and Tungodden, B., "You've Got Mail: A Randomized Control Field Experiment on Tax Evasion," Discussion Paper 26/2014, Department of Economics, Norwegian School of Economics, 2014.
- Boylan, S. J., and Sprinkle, G. B., "Experimental evidence on the relation between tax rates and compliance: The effect of earned vs. endowed income," *Journal of the American Taxation Association*, 23(1), 2001, pp. 75-90.
- Braithwaite, V., "Tax evasion," In *The Oxford Handbook of Crime and Public Policy* edited by Tonry, M., Oxford, Oxford University Press, 2011, pp. 381-406.
- Buckmann and Joseph, "An interpretable machine learning workflow with an application to economic forecasting," Staff Working Paper, No. 984, 2022.
- Carroll, J., "How taxpayers think about their taxes: frames and values," *Why people pay taxes: tax compliance and enforcement*, edited by Slemrod, J., Ann Arbor, MI: University of Michigan Press, 1992, pp. 43-63.
- Christian, C. W., "Voluntary compliance with the individual income tax: results from the 1988 TCMP study," *The IRS Research Bulletin*, 1500 (9-94), 1994, pp. 35-42.
- Clotfelter, C., "Tax evasion and tax rates: an analysis of individual returns," *The Review of Economics and Statistics*, 65(3), 1983, pp. 363-373.
- Collins, J. H., Milliron, V. C., and Toy, D. R., "Determinants of tax compliance: A contingency approach," *The Journal of the American Taxation Association*, 14(Fall), 1992, pp. 1-29.
- Cummings, R. G., Martinez-Vazquez, J., McKee, M., and Torgler, B., "Tax morale affects tax compliance: Evidence from surveys and an

- artefactual field experiment,” *Journal of Economic Behavior & Organization*, 70(3), 2009, pp. 447~457.
- Daude, C., Gutiérrez, H., and Melguizo, Á., “What drives tax morale?,” OECD DEVELOPMENT CENTRE, Working Paper No. 315, 2012.
- De Juan, A., Lasheras, M. A., and Mayo, R., “Voluntary compliance and behavior of spanish taxpayers,” Instituto de Estudios Fiscales, Madrid, Spain, 1993.
- Devlin, J., Chang, M., Lee, K., and Toutanova, K., “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” arXiv: 1810.04805, 2018.
- Dubin, J. A. and Wilde, L. L., “An empirical analysis of federal income tax auditing and compliance. *National Tax Journal*, 41(1), 1988, pp. 61~74.
- Dwenger, N., Kleven, H., Rasul, I., and Rincke, J., “Extrinsic and intrinsic motivations for tax compliance: Evidence from a field experiment in Germany,” *American Economic Journal: Economic Policy*, 8(3), 2016, pp. 203~32.
- Elster, J., “Social norms and economic theory,” *Journal of economic perspectives*, 3(4), 1989, pp. 99~117.
- Fehr, E. and Gächter, S., “Reciprocity and economics: The economic implications of homo reciprocans,” *European Economic Review*, 42(3-5), 1998, pp. 845~859.
- Feinstein, J. S., “An econometric analysis of income tax evasion and its detection,” *The RAND Journal of Economics*, 22(1), 1991, pp. 14~35.
- Feld, L. P. and Kirchgässner, G., “Direct democracy, political culture, and the outcome of economic policy: a report on the Swiss experience,” *European Journal of Political Economy*, 16(2), 2000, pp. 287~306.
- Fellner, G., Sausgruber, R., and Traxler, C., “Testing enforcement strategies in the field: Threat, moral appeal and social information,” *Journal*

- of the European Economic Association*, 11(3), 2013. pp. 634~660.
- Fischer, C. M., Wartick, M., and Mark, M., "Detection probability and taxpayer compliance: A review of the literature," *Journal of accounting literature*, 11, 1992, pp. 1~46.
- Fishlow, A., and Friedman, J., "Tax evasion, inflation and stabilization," *Journal of Development Economics*, 43(1), 1994, pp. 105~123.
- Frey, B. S., "Not Just for the Money. An Economic Theory of Personal Motivation," Cheltenham, UK: Edward Elgar Publishing, 1997.
- Frey, B. S. and Eichenberger, R., "The New Democratic Federalism for Europe," Cheltenham, UK: Edward Elgar Publishing, 1999.
- Friedland, N., "A note on tax evasion as a function of the quality of information about the magnitude and credibility of threatened fines: some preliminary research," *Journal of Applied Social Psychology*, 12(1), 1982, pp. 54~59.
- Friedland, N., Maital, S., and Rutenberg, A., "A simulation study of income tax evasion," *Journal of Public Economics*, 10(1), 1978, pp. 107~116.
- Gentzkow, M., and Shapiro, J. M., "What drives media slant? Evidence from US daily newspapers," *Econometrica*, 78(1), 2010, pp. 35~71.
- Gentzkow, M., Kelly, B., and Taddy, M., "Text as data," *Journal of Economic Literature*, 57(3), 2019, pp. 535~574.
- Gërxhani, K. and Schram, A., "Tax evasion and income source: A comparative experimental study," *Journal of Economic Psychology*, 27(3), 2006, pp. 402~422.
- Groseclose, T. and Milyo, J., "A measure of media bias," *The Quarterly Journal of Economics*, 120(4), 2005, pp. 1191~1237.
- Guala, F. and Mittone, L., "Experiments in economics: External validity and the robustness of phenomena," *Journal of Economic Methodology*, 12(4), 2005, pp. 495~515.

- Kidder, R. and McEwen, C., "Taxpaying behavior in social context: a tentative typology of tax compliance and noncompliance," In J. A. Roth and J. T. Scholz (eds), *Taxpayer Compliance*, Vol. 2 (pp. 46~75). Philadelphia: University of Pennsylvania, 1989.
- Kim, C. K., "Does fairness matter in tax reporting behavior?," *Journal of Economic Psychology*, 23(6), 2002, pp. 771~785.
- Kirchler, E., "The economic psychology of tax behaviour," Cambridge University Press, 2007.
- Kirchler, E. and Hölzl, E., "Twenty-five years of the *Journal of Economic Psychology* (1981-2005): A report on the development of an interdisciplinary field of research," *Journal of Economic Psychology*, 27(6), 2006, pp. 793~804.
- Kirchler, E., Muehlbacher, S., Kastlunger, B., and Wahl, I., "Why pay taxes? A review of tax compliance decisions," *Developing Alternative Frameworks for Explaining Tax Compliance*, 2010, pp. 1~27.
- Knack, S. and Keefer, P., "Does social capital have an economic payoff? A cross-country investigation," *Quarterly Journal of Economics*, 112(4), 1997, pp. 1251~1288.
- Lang, O., Nöhrbaß, K.-H., and Stahl, K., "On income tax avoidance: the case of Germany," *Journal of Public Economics*, 66(2), 1997, pp. 327~347.
- Lewis, A., "An empirical assessment of tax mentality," *Public Finance= Finances Publiques*, 34(2), 1979, pp. 245~257.
- Mason, R. and Calvin, L. D., "A study of admitted income tax evasion," *Law & Soc'y Rev.*, 13, 1978, pp. 74~89.
- Mittone, L., "Subjective versus objective probability: results from seven experiments on fiscal evasion," 1997.
- _____, "Dynamic behaviour in tax evasion: An experimental approach," *The Journal of Socio-Economics*, 35(5), 2006, pp. 813~835.

- Moser, D. V., Evans III, J. H., and Kim, C. K., "The effects of horizontal and exchange inequity on tax reporting decisions," *Accounting Review*, 70(4), 1995, pp. 619~634.
- Murphy, K., "The role of trust in nurturing compliance: A study of accused tax avoiders," *Law and Human Behavior*, 28(2), 2004, pp. 187~209.
- OECD, "Understanding and influencing taxpayers' compliance behaviour," In Forum on Tax Administration: Small/Medium Enterprise Compliance Subgroup, 2010, p. 40.
- Park, C. G. and Hyun, J. K., "Examining the determinants of tax compliance by experimental data: A case of Korea," *Journal of Policy Modeling*, 25(8), 2003, pp. 673~684.
- Pommerehne, W. W. and Weck-Hannemann, H., "Tax rates, tax administration and income tax evasion in Switzerland," *Public Choice*, 88(1), 1996, pp. 161~170.
- Pommerehne, W. W., Hart, A., and Frey, B. S., "Tax morale, tax evasion and the choice of policy instruments in different political systems," *Public Finance= Finances publiques*, 49(Supplement), 1994, pp. 52~69.
- Porcano, T. M., "Correlates of tax evasion," *Journal of Economic Psychology*, 9(1), 1988, pp. 47~67.
- Richardson, G., "The impact of tax fairness dimensions on tax compliance behavior in an Asian jurisdiction: The case of Hong Kong," *International Tax Journal*, 32, 2006, pp. 29~42.
- Röder, M., Both, A., and Hinneburg, A., "Exploring the space of topic coherence measures," In Proceedings of the eighth ACM international conference on Web search and data mining, 2015, February, pp. 399~408.
- Sandford, C. (Ed.), "Tax compliance costs measurement and policy,"

- Fiscal, 1995.
- Sandmo, A., "The theory of tax evasion: A retrospective view," *National tax journal*, 58(4), 2005, pp. 643~663.
- Schwartz, R. D. and Orleans, S., "On legal sanctions," *The University of Chicago Law Review*, 34(2), 1967, pp. 274~300.
- Slemrod, J., "An empirical test for tax evasion," *The review of Economics and Statistics*, 67(2), 1985, pp. 232~238.
- _____, "Methodological issues in measuring and interpreting taxable income elasticities," *National Tax Journal*, 51(4), 1998, pp. 773~788.
- Slemrod, J., Blumenthal, M., and Christian, C., "Taxpayer response to an increased probability of audit: evidence from a controlled experiment in Minnesota," *Journal of public economics*, 79(3), 2001, pp. 455~483.
- Spicer, M. W. and Becker, L. A., "Fiscal inequity and tax evasion: An experimental approach," *National Tax Journal*, 33(2), 1980, pp. 171~175.
- Spicer, M. W. and Hero, R. E., "Tax evasion and heuristics: A research note," *Journal of Public Economics*, 26(2), 1985, pp. 263~267.
- Spicer, M. W. and Lundstedt, S. B., "Understanding tax evasion," *Public Finance= Finances Publiques*, 31(2), 1976, pp. 295~305.
- Srinivasan, T. N. "Tax evasion: A model," *Journal of Public Economics*, 2(4), 1973, pp. 339~346.
- Tauchen, H. V., Witte, A. D., and Beron, K. J., "Tax compliance: An investigation using individual taxpayer compliance measurement program (TCMP) data," *Journal of Quantitative Criminology*, 9(2), 1993, pp. 177~202.
- Tittle, C. R. and Logan, C. H., "Sanctions and deviance: Evidence and remaining questions," *Law & Society Review*, 7(3), 1973, pp. 371~392.
- Torgler, B., "Speaking to theorists and searching for facts: Tax morale and

- tax compliance in experiments,” *Journal of Economic Surveys*, 16(5), 2002, pp. 657~683.
- _____, “Tax morale: Theory and empirical analysis of tax compliance (Doctoral dissertation, University_of_Basel),” 2003.
- _____, “Moral suasion: An alternative tax policy strategy? Evidence from a controlled field experiment in Switzerland,” *Economics of Governance*, 5(3), 2004, pp. 235~253.
- _____, “Tax morale and compliance: review of evidence and case studies for Europe,” World Bank Policy Research Working Paper, (5922), 2011.
- Torgler, B. and Murphy, K., “Tax morale in Australia: what shapes it and has it changed over time?,” *Journal of Australian taxation*, 7(2), 2004, pp. 298~335.
- Trivedi, V. U., Shehata, M., and Lynn, B., “Impact of personal and situational factors on taxpayer compliance: An experimental analysis,” *Journal of Business Ethics*, 47(3), 2003, pp. 175~197.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... and Polosukhin, I., “Attention is all you need,” *Advances in neural information processing systems*, 30, 2017.
- Vogel, J., “Taxation and public opinion in Sweden: an interpretation of recent survey data,” *National Tax Journal*, 27(4), 1974, pp. 499~513.
- Wärneryd, K. E., and Walerud, B., “Taxes and economic behavior: Some interview data on tax evasion in Sweden,” *Journal of Economic Psychology*, 2(3), 1982, pp. 187~211.
- Webley, P., Morris, I., and Amstutz, F., “Tax evasion during a small business simulation,” In H. Brandstatter and E. Kirchler (eds), *Economic Psychology*, Linz: Trauner, 1985, pp. 233~242.
- Webley, P., Robben, H., Elffers, H., and Hessing, D., “Tax Evasion: An

- Experimental Approach,” Cambridge: Cambridge University Press, 1991.
- Weck-Hannemann, H., and Pommerehne, W. W., “Einkommenssteuerhinterziehung in der Schweiz: Eine empirische Analyse,” *Schweizerische Zeitschrift für Volkswirtschaft und Statistik*, 125(4), 1989, pp. 515~556.
- Wenzel, M., “An analysis of norm processes in tax compliance,” *Journal of Economic Psychology*, 25(2), 2004, pp. 213~228.
- Yankelovich, Skelly and White, Inc., “Taxpayer attitudes survey: final report, public opinion survey prepared for the Public Affairs Division,” Internal Revenue Service, New York, 1984.
- Yitzhaki, S., “A note on income tax evasion: a theoretical analysis,” *Journal of Public Economics*, 3(2), 1974, pp. 201~202.
- Zhao, W. X., Jiang, J., Weng, J., He, J., Lim, E. P., Yan, H., and Li, X., “Comparing twitter and traditional media using topic models,” In *European Conference on Information Retrieval*, Springer, Berlin, Heidelberg, 2011, April, pp. 338~349.

〈웹사이트〉

- Ranks NL company, 「Korean Stopwords」, <https://www.ranks.nl/stopwords/korean>, 검색일자: 2022. 10. 4.

부 록

1. 국회예산정책처 보고서의 세법개정안 쟁점 모음

© 2013년

세목	쟁점	내용	세수효과	일몰
소득세	소득세 공제방식의 변경	근로소득공제를 축소 및 자녀양육 관련 인적공제 및 특별공제의 세액공제 전환, 소득공제에서 세액공제 형태로 변화	2015년 11,388억원	
	근로장려세제 확대 및 자녀장려세제 신설	기초생활수급자 포함, 단독가구 적용 확대, 지급기준 변경 및 소득 기준 완화, 자녀장려세제 도입, 재산요건 완화, 주택요건 완화	2014년 -4,315억원	
	신용카드 공제를 인하	공제를 하향 조정(15 → 10%)	2014년 -569억원	
	월세소득공제 확대	무주택 세대원 공제 하향, 공제를 및 공제한도 인상(50 → 60%, 300만 → 500만원)	연간 -1,3억원	
	엔젤투자 소득공제 확대	공제율 인상, 공제한도 확대, 소득공제 종합한도 초과금액 5년간 이월공제 하향	(2014년~2018년) 총 -87억원	
	증권인과세	2015년부터 증권인의 소득에 대해 과세		

세목	쟁점	내용	세수효과	일몰
법인세	R&D 부문 세액공제	적용대상 확대	2014년 -801억원	
	기타 투자지원 제도	에너지절약시설, 환경보전시설, 의약품품질관리 개선시설 투자세액공제 일몰연장 및 공제율 조정	2014년 -418억원	
	중소기업특별세액감면제도	대상업종 확대	추정 곤란	
부가가치세	일자리 창출을 위한 세제지원	고용창출투자 세액공제, 고용중기인원 사회보험료 세제지원 등		
	면세농수산물 등의 의제매입세액공제 한도설정 등	매출액의 30% 공제한도 제한	2014년 2,584억원	
	재활용폐자원 등 의제매입세액공제를 조정	의제매입세액공제를 3/103(중고자동차 5/105) 인하	(2014년~2016년) 총 1조 1,005억원	2014년 일몰
재산세	치료 이외 미용·성형용역의 면세대상 제외	의료보건용역의 과세범위 확대		
	일감몰아주기 증여세 과세제도 완화	중소기업 과세요건 완화, 특수법인 간 내부거래 과세제외 범위 확대, 소득세 이중과세 조정		
	상중세법상 기업승계 적용대상 확대 및 사후관리 요건 완화	매출액 3,000억원 이하 중견기업까지 확대, 유류분 반환청구 공동 상속 인정 등		
총 장점 수	15개			

자료: 국회예산정책처, 『2013~2017년 세법개정안 분석』, 『2018~2020년 시행 개정세법의 주요 내용 및 심사 쟁점』, 『2020~2021 개정세법의 심사 경과와 주요 내용』을 참고하여 저자 작성

© 2014년

세목	정점	내용	세수효과	일몰
소득세	배당소득 증대세제 신설 비과세종합저축 등 과세특례 저축상품 통합 재설계	배당 상황 증가 시 배당소득세를 인하 고령자 연령 상향(60 → 65세), 저축한도 확대(3천만 → 5천만원), 비과세	매년 -224억원 (2016년~2019년) 총 -10,033억원	
	퇴직연금 납입한도 및 퇴직소득 과세제도 개선	퇴직연금 추가 300만원 공제한도 허용, 퇴직소득 차등공제 도입, 과세 표준 상향 조정	① 공제확대 (연평균 - 2,933억원) ② 차등과세 총 2,233억원 증가	
법인세	기업소득 환류세제	투자외 배당 인금증가 등이 소득 일정액에 미달 되는 경우 추가 법인세 과세		
	근로소득 증대세제	임금증가분의 일정률을 소득세액과 법인세액에서 세액공제	2016년 -2,231억원	
	고용창출투자세액공제	고용유지 혹은 증가된 경우 공제, 총 공제율(4~7% → 4~9%)	2015년 549억원	2017.12.31
	조합법인 등 법인세 과세특례	일몰 3년 연장, 융합 수협 등은 단일세율 9% 적용	2015년 422억원	항후 3년
	간접의국납주세액공제 기준 강화	국외 자회사 지분을 강화(10 → 25%), 국외순회사 제외, 국별한도액 계산 방식만 인정	매년 1,040억원	
부가가치세	금융 보험업외의 부가가치세 면세범위 축소 해외 오픈마켓에서 구매하는 전자적 용역 과세 중고자동차에 대한 매입세액공제 축소	부수적 금융서비스에 대해 과세 전환 해외개발자 앱 등도 해외 오픈마켓 사업자를 통해 부가가치세 과세 2015~2016년 7/107, 2017년 이후 5/105	연평균 1조9,432억원 (2015년~2019년) 약 7,347억원	
	소비세	담배가격 인상에 따른 세제 개편	(부가가치세) (2015년~2019년) 총 21,741억원	
재산세	「상속세 및 증여세법」상 기업승계 적용대상 확대 및 사후관리 요건 완화	매출액 5,000억원 미만까지 확대, 요건 및 사후관리무 대폭 완화		
총 정점 수	13개			

자료: 국회예산정책처 「2013~2017년 세법개정안 분석」, 「2018~2020년 시행 개정세법의 주요 내용 및 심사 정점」, 「2020~2021 개정세법의 심사 경과와 주요 내용을 참고하여 저자 작성

© 2015년

세목	쟁점	내용	세수효과	일몰
소득세	비과세 개인종합자산관리계좌 도입	ISA 제도 신설, 근로소득자 및 개인 사업자로 대상 확대, 직전 연도 2천만원 이상 금융소득 신고자는 대상 제외	(연간 - 1,100억원~ -5,500억원)	
	해외주식투자저용펀드 도입	1인당 최대 납입금액 3천만원까지 10년 동안 양도차익, 환차익 비과세	추정 곤란	
	상장주식 양도소득세 대주주 범위 확대 및 중소기업 대주주 세율 인상	유기증권시장 50억 → 25억원 코스닥시장 40억 → 20억원으로 축소 하여 과세대상을 확대, 중소기업 대주주 세율 10 → 20%	추정 곤란	
	종교소득 과세	2013년 시행령 근거 마련, 2014년 시행령 개정으로 1년 유예, 종교 소득 신설하여 법률 정비		
법인세	청년고용증대세제	청년 상시근로자 증가 인원당 대기업 250만원, 중견, 중소기업 500만원 공제		향후 3년
	R&D 부문 세액공제	3년 연장, 인건비 대상 연구진담요원 연구보조원 한정, R&D 설비 투자 세액공제를 대기업(3 → 1%), 중견기업(5 → 3%), 중소기업(10 → 6%) 인화	(2016~2020년) 1조 8,712억원	2018.12.31
	법인 등 업무용 승용차 관련 비용 과세 합리화	필요경비 및 손금 인정 기준 명확히 함	연평균 1,100억원	
소비세	부가가치세의 매입자 납부특례 대상 확대	매입자 납부방식으로 전환		
	개별소비세 과세대상 및 기준가격, 세율 조정	보석류 등 과세기준 200만 → 500만원, 고급가구 개당 1,000만원/조당 1,500만원, 에너지다소비 가전제품 과세대상 폐지	향후 5년간 -6,950억원	
총 쟁점 수	10개	증여 및 증여재산의 개념과 과세대상 증여재산의 범위 명확히 함		

자료: 국회예산정책처, 『2013~2017년 세법개정안 분석』, 『2018~2020년 시행 개정세법의 주요 내용 및 심사 쟁점』, 『2020~2021 개정세법의 심사 경과와 주요 내용』을 참고하여 저자 작성

© 2016년

세목	쟁점	내용	세수효과	일몰
소득세	신용카드 등 사용액 소득 공제연장 및 한도조정	3년 연장, 공제금액 한도 차등 적용	(2018~2021년) 총 2,398억원	
	근로장려금 신청액 상향	10% 인상	매년 -950억원	
	출생·인양 세액공제 확대	통제 이상 출산 시 세액공제 확대	매년 -260억원	
	등본학자금 대출 등의 원리금상환액에 대해 교육비 세액공제 적용	공제율 15%	(2018~2021년) 총 -3,420억원	
	외국인근로자 과세특례 적용기한 연장 및 세율조정	17 → 19%	2017년 472억원	
법인세	국내주식 양도소득세 예정 신고 횟수 조정	분기별 1회 → 반기별 1회로 축소	2018년 -4,891억원	
	영상콘텐츠 제작비용에 대한 세액공제 신설	중견, 대기업: 20%+a, 중소기업: 30%	(2018~2020년) 총 -1,147억원	
	고용·투자·R&D 관련 세제지원 대상업종 확대		2017년 -200억원	
	중소기업 고용창출 투자세액공제 한도 확대		2017년 -85억원	
	내국법인인 벤처기업 등에 출자 시 세액공제 신설		(2018~2020년) 총 -630억원	
	중소기업 기술실험특례 적용기한 연장 및 대상 확대		매년 -211억원	
	안정성비 등 투자세액공제 대상에 내진보강 설비 추가		2017년 -58억원	
	액상형 분류에 대한 부가기치세 면제		2017년 -18억원	
	가산세 부담 완화	납세협력의무 위반에 대한 가산세 부담 50% 경감	2017년 -234억원	
	발전용 유연탄에 대한 개별소비세율 조정	1kg당 30원으로 상향	2017년 3,565억원	
총 쟁점 수	15개			

자료: 국회예산정책처, 『2013~2017년 세법개정안 분석』, 『2018~2020년 시행 개정세법의 주요 내용 및 심사 쟁점』, 『2020~2021 개정세법의 심사 경과와 주요 내용』을 참고하여 저자 작성

© 2017년

세목	쟁점	내용	세수효과	일몰
소득세	소득세를 인상안	과표 3억~5억원: 38 → 40% / 5억원 초과: 40 → 42%	(연평균) 9,761억원	
	대주주 주식양도차익 과세 강화	① 대주주 주식양도차익 과표 3억원 초과 구간 세율 5%p인상(20 → 25%) ② 대주주 범위 확대(2018년 4월 종목별 보유액 15억원 이상, 2020년 4월 10억원 이상 → 2021년 4월 3억원 이상) * 2020년 세법 시행령 개정안에서 10억원 유지	(연평균) 3,318억원	
	다주택자 양도소득세 인상 등	조정대상지역 내 1세대 2주택 이상 보유자의 양도소득세율 인상 및 장기보유특별공제 배제	추정 곤란	
	근로장려금 지급 확대 산업액 상향조정	(약 10%): 77,185,230만원 → 85,200,250만원 및 취약계층 지원 확대	(연평균) -1,296억원	
	종교인소득이 있는 거주자에 대한 근로·자녀장려금 신청 허용	종교인소득의 신고·납부방식에 관계없이 근로·자녀장려금의 신청 허용	추정 곤란	
	월세 세액공제를 인상	총급여 5,500만원 이하 무주택 세대주 월세액 공제를 10 → 12%	(연평균) -69억원	
	개인종합자산관리계좌(SA) 제도 개선	비과세 한도 인상(200만~250만원→200만~400만원) 및 중도인출 허용	추정 곤란	2018.12.31.
	자녀세액공제 축소·정비	6세 이하 공제 2018년부터 폐지, 만 0~5세 아동 기본공제(1명당 15만원, 셋째부터 30만원)는 1년 유예 후 2019년 폐지	(연평균) 2,740억원	
	장기보유특별공제 합리화	보유기간 3~15년 이상, 공제율 6~30%(연간 2%, 최대 15년)	(연평균) 3,267억원	
	벤처기업 등에 대한 세제지원 확대	벤처기업 투자에 대한 과세특례 연장	(연평균) -1억원	2020.12.31.
법인세	법인세율 인상	3,000억원 초과 구간을 신설, 25%의 세율을 적용 * 정부안 2,000억원	(연평균) 21,111억원	
	고용증대세제(청년고용증대세제)	투자가 없더라도 고용 증가기원 1인당 일정금액 공제(중소·중견기업은 2년간), 다른 고용·투자지원 제도와 중복 허용	(연평균) -6,355억원	
	근로소득 증대세제 확대	세액공제 적용기간 3년 연장, 공제율 10→20%	(연평균) -18억원	2020.12.31.

세목	쟁점	내용	세수효과	일몰
법인세	정규직 근로자로의 전환에 따른 세액공제 확대	적용기한 1년 연장, 전환인원 1명당 공제액 상향(중소: 700만→1천만 원, 중견: 500만→700만원)	(연평균) -41억원	2018.12.31.
	중소기업 고용증가 인원에 대한 사회보험료 세액공제 확대 등	공제기간 2년으로 확대, 사회보험 신규가입자에 대해 중소기업 부담 사회보험료 세액공제 신설	(연평균) -1,177억원	2018.12.31.
	투자·상생협력촉진세제 신설	기업소득환류세제의 환류대상 및 기준치 등을 조정한 투자·상생협력 촉진세제로 전환	추정 곤란	2020.12.31.
	대기업 연구인력개발비 세액공제 축소	당기분: -1%p, 증기분: -5%p	(연평균) 1,069억원	
	신성장동력·원천기술 연구개발비 세액공제 확대	최대 30→40%	(연평균) -114억원	2018.12.31.
	설비투자세액공제 축소	생산성 향상시설 등 투자세액공제 연장 및 대기업/중견기업 공제율 축소	(연평균) 3,862억원	2019.12.31
	중소기업 특별세액 감면 재실계	적용기한 3년 연장 및 감면한도 1억원 신설, 중복적용 허용	(연평균) 2,391억원	2020.12.31.
	조합법인 과세특례 일몰연장 등	일몰기한 3년 연장	세수변동 없음	2020.12.31.
	경력단절여성 재고용 중소기업에 대한 세액 공제 확대	중견기업까지 적용 확대, 공제율 인상, 일몰기한 3년 연장	추정 곤란	2020.12.31.
	상생결재 지급금액에 대한 세액공제 확대	중견기업까지 적용 확대, 일몰기한 3년 연장	추정 곤란	2020.12.31.
소비세	면세농산물 등 의제매입세액공제율 인상 및 공제율을 법률로 상향	과세표준 2억원 이하인 개인 음식점 사업자의 경우 2년간 일시적으로 공제율을 8/108 → 9/109로 상향 조정	585억원 (연평균) -741억원	2년간
	수정수입세금계산서 발급시유 확대	수입자의 경미한 과실인 경우가 입증되는 경우에도 세금계산서 발급	추정곤란	
	발전유연단 개별소비세율 인상	기본세율 인상(kg당, 30→36원)	(연평균) 4,885억원	
	종교자동차 의제매입세액공제율 상향	9/109→10/110으로 상향 조정	(연평균) -351억원	2018.12.31
	농·어업용 기자재 등 공급에 대한 부가가치세 영세를 적용기한 연장	적용기한 2년 연장	세수변동 없음	2020.12.31.
	농협이 공급하는 전산용역 부가가치세 면제	지회사 공급 전산용역에 대해 부가가치세 면제	세수변동 없음	2020.12.31.

세목	쟁점	내용	세수효과	입몰
소비세	전자자동차에 대한 개별소비세 전면 일몰연장	3년 연장	세수변동 없음	2020.12.31.
	영·유아용 기저귀·분유에 대한 부가가치세 면제 일몰기한 연장	3년 연장	세수변동 없음	2020.12.31.
	일반택시 회사의 부가가치세 납부세액 경감을 확대	부가가치세 납부세액 경감을 95→99%	(연평균) -96억원	2018.12.31.
	국민주탁규모 초과 공동주택 관리용역 등에 대한 부가가치세 면제 일몰기한 연장	3년 연장	세수변동 없음	2020.12.31.
	외국인 관광객 숙박요금에 대한 부가가치세 환급 제도 제도인	30일 이하의 관광호텔 숙박용역에 대한 부가가치세 환급	추정 곤란	2018.12.31.
	외국인 관광객 미용성형 의료용역에 대한 부가가치세 환급특례 일몰기한 연장	1년 연장	세수변동 없음	2018.12.31.
	신용카드사 부가가치세 대리납부제도 신설	채납률 높은 특정 업종의 부가가치세 대리납부 및 세액공제(1%) 허용 126 → 529원 * 국제 개정(2017. 11. 9.)	(연평균) 210억원 추정 곤란	2021.12.31.
	관련항 전자담배 개별소비세 인상	대기업의 과세대상 확대, 대·중견기업의 증여제이의 신장기준 확대 및 과세 제외 중소기업의 범위 조정	(연평균) 545억원	
	기업상속공제 공제조건 강화	기업상속공제의 공제한도별 기업영위기간 상향 조정		
	상속·증여 신고세액공제 축소	현행 7%에서 2018년 5%, 2019년 이후 3%로 단계적 인하	450억원 (연평균) 1,210억원	
지방세	관련항 전자담배 세율인상에 따른 담배소비세율 조정	추정 곤란		
총 쟁점 수	41개			

자료: 국회예산정책처, 「2013~2017년 세법개정안 분석」, 「2018~2020년 시행 개정세법의 주요 내용 및 심사 쟁점」, 「2020~2021 개정세법의 심사 경과와 주요 내용」을 참고하여 저자 작성

© 2018년

세목	정책	내용	세수효과	일몰
소득세	근로장려세제(EITC) 확대	연령요건 폐지 및 소득·재산요건 완화, 지급액 인상, 지급시기 변경 등	(연평균) -22,851억원	
	주택 임대소득 과세 개편	기본공제 및 필요경비율 차등 적용, 전세 임대보증금 과세 배제, 소형 주택기준 축소, 분리과세 주택임대소득만 있는 사업자 추가 * 2014년 개정사항에 포함된 내용, 2년간 과세 유예기간 부여, 재수정에 따라 3년간 유예기간, 2016년 세법개정에 따라 2019년으로 연기	(연평균) 257억원	
	신용카드 등 소득공제 일몰연장	일몰 1년 연장, 박물관, 미술관 입장료 추가공제 신설	세수변동 없음, 추가 공제에 대한 세수효과는 추정 곤란	2019.12.31.
	상호금융기관 예탁금·출자금 과세특례 일몰연장	2년 연장	세수변동 없음	
법인세	기부금 세액공제 확대	기부금 고액기준 하향(2천만 → 1천만원) 및 초과금액 이월공제 기간 연장(5 → 10년)	(연평균) -99억원	
	아동수당 대상 확대에 따른 자녀세액공제(1인당 15만원 공제) 조정	자녀세액공제 중 1인당 15만원 공제 대상 조정(만 6세 이상 자녀 → 만 7세 이상 자녀)	(연평균) 619억원	
	특정 설비·시설투자 세액공제제도 공제율 정비 및 일몰연장	특정 설비·시설투자 세액공제 공제율 정비 및 일몰연장, 안전설비투자 세액공제 대상시설 추가강화된 안전기준에 따른 소방시설)	(연평균) -172억원	2021.12.31.
	신성장동력·원천기술 R&D비용 및 시설투자 세제지원 확대	신성장 R&D 비용 세액공제 대상 확대 및 일몰연장, 신성장기술 사업 회사설 투자세액공제 요건 완화 및 일몰연장, 초연결 네트워크 구축 설비투자 세액공제 신설	세수변동 없음, 추정 곤란	2020.12.31.
법인세	고용증대세제 확대	적용기한 확대 및 청년 정규직 고용 시 1인당 감면액 100만원 추가	(연평균) -1,968억원	2021.12.31.
	위기지역 지원세제 신설	고용위기지역 또는 산업위기대응특별지역에 창업한 기업에 대한 세액감면 및 투자세액 공제를 상향	(연평균) -60억원	

세목	정책	내용	세수효과	일몰
법인세	지역특구 조세감면 제도 일몰연장	3년 연장	추정 곤란	2021.12.31
	해외진출기업의 국내복구에 대한 세액공제 확대 연장	국내 부분복구 공제, 지방 부분복구에 대한 공제를 상향	세수변동 없음	2021.12.31.
	부가가치세액 중 지방소비세 비율 인상	11 → 15%	(연평균) -35,633억원	
	제조업 중 도정업 등 개인사업자의 의제매입세액공제를 인상	도정업 등 개인사업자 공제율: 6/106	추정 곤란	
	부가가치세 납부면제점 인상	2,400만 → 3,000만원	(연평균) -236억원	
	국외사업자의 과세대상 전자적 용역 범위 확대	광고, 클라우드컴퓨팅서비스, 재화 용역 중개 과세대상 포함	추정 곤란	
	유연탄 세율 인상 및 발전용 천연가스 세율 인하	유연탄 세율(36원/kg → 46원/kg), 발전용 LNG 세율(60원/kg → 12원/kg)	(연평균) 1,016억원	
	신용카드 등의 사용에 따른 매출세액공제 확대공제를 적용기한 연장	3년 연장, 공제한도 3년 한시적 인상(500만 → 1천만원)	(연평균) -2,119억원	2021년
	농·임·어업용 및 연인여객선박용 석유류에 대한 부가가치세 등 감면 연장	농·임·어업용 3년 연장, 연인여객 2년 연장	세수변동 없음	농·임·어업용: 21.12.31. 연인여객선박용: 20.12.31.
	소비세	재활용폐자원 등에 대한 의제매입세액공제 적용기한 연장	재활용폐자원 3년, 종고자등차 1년 연장	세수변동 없음
노후경유자동차 교체에 대한 개별소비세 감면 신설		노후 경유차 폐차 후 신차 구매시 개소비세 70% 감면	(연평균) -983억원	
「교통·에너지·환경세법」 유효기간 연장		3년 연장	세수변동 없음	

세목	쟁점	내용	세수효과	일몰
	종합부동산세: 세율 인상	주택분 및 종합합산토지분의 세율을 각각 인상	(연평균) 12,229억원	
	종합부동산세: 공정시장이익비율 조정	매년 5%p씩 2022년 100%까지 단계적 인상	(연평균) 6,012억원	
	종합부동산세: 세부담 상한비율 인상	조정대상지역의 2주택 소유자, 3주택 이상 세부담 상한비율 인상 (150 → 200%, 300%)	추정 곤란	
	종합부동산세: 1세대 1주택에 대한 세액공제율 인상	15년 이상 장기보유한 1세대 1주택 공제율 인상	(연평균) -11억원	
재산세	종합부동산세: 분납대상 확대 및 분납기간 확대	분납대상(600만원 초과 → 250만원 초과), 분납기간(2개월 이내 → 6개월 이내)	(2019년) -239억원	
	상속세 및 증여세: 기업상속공제의 사후관리요건 완화 등	사후관리 의무 위반 시 자산처분비율을 고려	추정 곤란	
총 쟁점 수	28개			

자료: 국회예산정책처 『2013~2017년 세법개정안 분석』, 『2018~2020년 시행 개정세법의 주요 내용 및 심사 쟁점』, 『2020~2021 개정세법의 심사 결과와 주요 내용』을 참고하여 저자 작성

◎ 2019년

세목	쟁점	내용	세수효과	일몰
소득세	어로어업소득 비과세 확대	5천만원까지 비과세	추정 곤란	
	근로소득공제 한도 신설	2천만원 한도	(연평균) 754억원	
	개인종합자산관리계좌의 연금계좌 전환 시 연금계좌 세액공제 적용	납입액의 10%(300만원 한도)까지 세액공제	(연평균) -140억원	
	50세 이상 연금계좌세액공제 한도 확대	납입한도 연 600만원, 종합소득 1억원 초과자는 연 300만원	(연평균) -277억원	
	대토보상에 대한 양도소득세 감면율 인상	감면율 15 → 40%	(연평균) -37억원	
	사용금액에 대한 소득공제 적용기한 연장 등	3년 연장, 신문구독비용 추가공제	(연평균) -19억원	

세목	장점	내용	세수효과	일몰
법인세	생산성 향상시설 투자세액공제제도 공제를 정비 및 일몰연장	대기업 2%(1년 한시), 중견 5%, 중소기업 10%	(연평균) -1,932억원	
	영상콘텐츠 제작비용 세액공제 대상 확대 및 적용 기한 연장	오락 프로그램, 다큐로 적용대상 확대, 적용기간 3년 연장	(연평균) -366억원	
법인세	5G 이동통신 기지국 시설 투자세액공제 확대	기지국 시설투자비+공사비도 공제대상 포함	(연평균) -189억원	
	접대비 손금산입 한도 상향	중소기업 접대비 기본한도 2,400만 → 3,600만원 인상	(연평균) -1,777억원	
	경력단절여성 재고용 기업에 대한 세액공제 요건 완화	경력단절 기간 3~10년 이내 → 3~15년 이내	(연평균) -128억원	
	내국법인의 소재·부품·장비 전문 기업 출자·인수에 대한 과세특례	인수기업의 5%(중견기업 7%, 중소기업 10%) 세액공제 신설	추정 곤란	
	부가가치세의 지방소비세 이양비용 인상	15→21%	(연평균) -55,895억원 국가 전체의 세입 증감은 없음	
	면세농산물 등의 의제매입세액공제율 조정	공제를 인상 기한 2년 연장	세수변동 없음	
소비세	과세유형장소에 대한 공제를 인하	4/104→2/102	(연평균) 32억원	
	외국인 관광객에 대한 부가가치세 환급특례 일몰 기한 연장	숙박업 및 의료업 부가가치세 환급 특례 기한 1년 연장	세수변동 없음	
	농협이 공급하는 전산용역 부가가치세 면제	면제의 적용기한을 2년 연장	세수변동 없음	
	택수와 맥주의 과세체계 전환	중기세에서 증량세로 전환	(연평균) 14억원	
	등유 등을 차량연료로 사용 시 교통·에너지·환경세 부과	경유차량에 등유 등을 연료로 사용 시 경유세율 부과	(연평균) 16억원	
	수소전기차 개별소비세 감면 적용기한 연장	3년 연장	세수변동 없음	

세목	쟁점	내용	세수효과	일몰
소비세	노후차 교체 시 개별소비세 감면	개별소비세 70% 감면(6개월 한시)		
	증권거래세율 인하	0.5→0.45%	(연평균) -208억원	
재산세	가임상속공제 제도의 시후관리요건 완화	시후관리기간 10→7년	추정 곤란	
	최대주주 할증평가 제도 완화	최대주주 지분율에 따른 차등할증 폐지 및 중소기업 할증평가 적용 제외	추정 곤란	
	외부감사대상성의 확대 및 감사인 지정 제도·회계 관리제도 도입	외부감사 대상기준 수입금액 추가, 외부감사 공익법인 확대, 감사인 지정제도 도입	추정 곤란	
총 쟁점 수	25개			

자료: 국회예산정책처, 『2013~2017년 세법개정안 분석』, 『2018~2020년 시행 개정세법의 주요 내용 및 심사 쟁점』, 『2020~2021 개정세법의 심사 경과와 주요 내용』을 참고하여 저자 작성

◎ 2020년

세목	쟁점	내용	세수효과	일몰
소득세	금융투자소득 과세체계 도입	3억원 이하 20%, 3억원 초과 25% 세율 적용	(연평균) 9,808억원	
	소득세 최고세율 인상	10억원 초과 42 → 45%	(연평균) 9,645억원	
	개인종합자산관리계좌(SA) 제도 개선	가입대상 확대, 운용범위 확대, 의무가입기간 5→3년 완화, 이월납입 허용	추정 곤란	
	가상자산소득 기타소득세 신설	연 250만원 초과 가상자산소득에 대해 20% 분리과세 * 가상자산소득에 대한 과세 시행시기 유예	추정 곤란	
	특정사회기반시설 집합투자지구 투자자에 대한 과세특례 신설	뉴딜 인프라펀드 투자자 배당소득 2억원까지 9% 분리과세	추정 곤란	
장기일반민간임대주택 등에 대한 양도소득세 과세특례 축소	적용기한 2년 단축	추정 곤란		

세목	정책	내용	세수효과	일몰
법인세	통합투자세액공제 재설계	기본공제에 추가공제 3% 적용	(연평균) -4,267억원	
	기업환경 개선을 위한 세제 개편	이월공제기간 확대	(연평균) -4,801억원	
	지방이전기업 세액감면 제도 재설계 등	공장, 본사 이전 구분하여 세액감면 제도 운영	세수변동 없음	
	신탁재산 법인세 과세에 대한 특례	생신량 감축조건 폐지	추정 곤란	
	유턴기업에 대한 세액감면 확대	당기소득 적용 비율(60 → 70%) 상향, 환류 대상 범위 확대(7천만 → 8천만원), 초과환류액 이월기간 2년으로 확대		
	투자·상생협력촉진세제 일몰연장 및 재설계	일몰기한 2년 연장	세수변동 없음	
	조합법인 과세특례 일몰연장 등	배당소득 과세 * 미반영 정부안	추정 곤란	
	특정내국법인의 초과유보 소득 배당간주 신설 (정부안)	4,800만 → 8,000만원	(연평균) -2,245억원	
	간이과세 기준금액 인상 등	3,000만 → 4,800만원	(연평균) -2,245억원	
	남부면제 기준금액 인상	연초임이 아닌 다른 부분을 원료로 제조하는 것까지 과세	(연평균) 170억원	
소비세	과세대상 담배의 확대	「주세법」을 「주세법과 주류 면허 등에 관한 법률」로 분법		
	주세법 체계 개편 및 제도 정비	1m당 370→740원 * 미반영 정부안	(연평균) 33억원,	
	액상형 전자담배에 대한 개별소비세율 인상(정부안)	유연대용신탁 및 수익지연신탁 상속 정의 추가		
재산세	위탁자·수익자 사망과 연계한 새로운 유형의 신탁에 대한 과세의 명확화	수탁자에서 위탁자로 남세의무자 변경	(연평균) 516억원	
	신탁재산에 대한 종합부동산세 남세의무자 변경 등	일반공익법인과 성실공익법인을 공익법인으로 통합		
	공익법인 과세체계 개선 및 공익활동 강화 촉진	0.45 → 0.35%	(연평균) -469억원	
	증권거래세를 인화			

세목	쟁점	내용	세수효과	일몰
재산세	부동산 세제 개편(양도소득세, 법인세 포함, 종합 부동산세)	① 양도소득세: 주택 보유기간 1년 미만(40 → 70%), 2년 미만(기 본세율 → 60%) ② 법인세: 법인주택 양도소득 추가 과세율(10 → 20%) ③ 종합부동산세: 다주택자 보유과세 강화 * 의원안으로 발의 후 8월 4일 국회 본회의 통과		
총 쟁점 수	22개			

자료: 국회예산정책처, 『2013~2017년 세법개정안 분석』, 『2018~2020년 시행 개정세법의 주요 내용 및 심사 쟁점』, 『2020~2021 개정세법의 심사 경과와 주요 내용』을 참고하여 저자 작성

◎ 2021년

세목	쟁점	내용	세수효과	일몰
소득세	근로장려금 소득기준금액 인상	200만원 상향	(연평균) -2,628억원	
	상기임대료 인하한 임대사업자에 대한 세액공제 적용기간 연장	1년 연장		
	1세대 1주택 양도소득세 비과세 기준금액 상향	9억 → 12억원 * 정부안 미포함 의원안 발의	추정 곤란	
	가상자산소득 과세 시행 1년 유예	* 정부안 미포함 의원안 발의	추정 곤란	
	기부금 세액공제 한시 상향	세액공제 5%p 상향	(2022년) -3,223억원 2021년 한시	
	납입기술비 등 의로비 세액공제 확대	납입기술비 등 공제율 30%, 고위험임신부 등 공제율 20%	납입기술 공제를 확대: (연평균) -17억원	
	고향사랑 기부금에 대한 세액공제 신설	10만원 이하 90.9%, 10만원 초과 15%	추정 곤란	

세목	정점	내용	세수효과	일몰
법인세	국가전략기술 R&D·시설투자 공제율 신설 등 투자세액공제 확대	① 국가전략기술 R&D 및 시설투자 구간 신설 ② 신성장, 원전기술연구개발비 세액공제 일몰기한 3년 연장	① (연평균) -1조 988억원 ② (연평균) -6,836억원	2024.12.31.
	고용중대 세액공제 공제금액 한시적 상향 및 적용기한 연장	세액공제 적용기한 3년 연장 및 비수도권 취업 취약계층 우대공제금액 상향 (* 비수도권 우대공제 대상 공제금액 상향은 2021~2022년 한시)	(연평균) -1조 674억원	2024.12.31.
	창업 중소기업 세액감면 대상 확대 등	세액감면 적용기간 3년 연장, 연 수입금액 기준 4,800만원 이하 → 8천만원 이하	(연평균) -4,559억원	2024.12.31.
	중소기업 사회보험료 세액공제 실효성 제고 등	공제기간 동안 상시근로자 감소 시 잔여기간 공제 배제, 적용기간 3년 연장	(연평균) -2,934억원	2024.12.31.
	상생결제 지급금액에 대한 세액공제 지원 확대	한금성결제 비율 삭제, 공제율 상향(15일 이내 0.2 → 0.5%, 16~60일 0.1 → 0.3%, 30~60일(신설) 0.15%)	(연평균) -38억원	2022.12.31.
	부가가치세의 지방소비세 이양비용 인상	현행 21%에서 2022년 23.7%(이하 현행 대비+2.7%), 2023년 이후 25.3%(+4.3%p)로 인상	(연평균) -4.2조원	
	신용카드 매출세액공제 우대공제율 및 우대공제 한도 연장	공제율(1 → 1.3%) 및 공제한도(500만 → 1천만원)의 적용기한 2년 연장	(연평균) -4,413억원	
	농협 및 수협의 전산용역 부가가치세 면제 연장	적용기한 2년 연장	(연평균) -53억원	2023.12.31.
	교통·에너지·환경세 적용기한 연장 등	적용기한 3년 연장	세수변동 없음	2024.12.31.
	개별소비세 공제요건 완화	제조장 보세구역 이외 예외 인정	(연평균) -31억원	
위기지역·제주도 소재 골포장 개별소비세 감면 종료				

세목	쟁점	내용	세수효과	일몰
재산세	가업상속공제 대상기업 범위 확대	연매출 3천억 → 4천억원 미만	추정 곤란	
	영농상속공제 한도 상향 등	15억 → 20억	추정 곤란	
	상속세 물납 대상 확대	문화재, 미술품 물납 허용	세수효과 없음	
	1세대 1주택자에 대한 주택분 종합부동산세 기 분공제액 상향	9억→11억원 * 정기국회 이전에 의결된 「종합부동산세법」 개정안 발의일자: 2021. 8. 25. 의결일자: 2021. 8. 31.	(연평균) -836억원	
총 쟁점 수	22개			

자료: 국회예산정책처, 「2013~2017년 세법개정안 분석」, 「2018~2020년 시행 개정세법의 주요 내용 및 시사 쟁점」, 「2020~2021 개정세법의 심사 경과와 주요 내용을 참고하여 저자 작성

2. 설문조사 예시

1-5회차 설문조사

1. (가상자산소득 과세) 정부는 2023년부터 가상자산에서 발생한 양도차익에 대해 20%의 세율을 적용하여 과세할 예정입니다. 귀하는 가상자산 양도차익에 대해서 과세하는 것에 대해 어떠한 입장이십니까?

- ① 매우 긍정적이다
- ② 대체로 긍정적이다
- ③ 보통이다
- ④ 대체로 부정적이다
- ⑤ 매우 부정적이다

2. (금융투자소득세) 정부는 2023년부터 주식 거래에서 발생한 양도차익에 대해 20%의 세율을 적용하여 과세할 예정입니다. 귀하는 주식 거래 양도차익에 대해서 과세하는 것에 대해 어떠한 입장이십니까?

- ① 매우 긍정적이다
- ② 대체로 긍정적이다
- ③ 보통이다
- ④ 대체로 부정적이다
- ⑤ 매우 부정적이다

3. (종합부동산세) 정부는 재산세 과세대상인 주택 및 토지를 유형별로 구분하여 인별로 합산한 결과 그 공시가격 합계액이 유형별 공제금액을 초과하는 경우 그 초과분에 대해서 종합부동산세를 과세하고 있습니다. 귀하는 종합부동산세를 과세하는 것에 대해 어떠한 입장이십니까?

- ① 매우 긍정적이다
- ② 대체로 긍정적이다
- ③ 보통이다
- ④ 대체로 부정적이다
- ⑤ 매우 부정적이다

4. (근로소득세) 정부는 고용주에게 근로를 제공하고 그 대가로 지급받은 소득에 대해서 근로 소득세를 과세하고 있습니다. 귀하는 근로 소득에 대하여 과세하는 것에 대해 어떠한 입장이십니까?

- ① 매우 긍정적이다
- ② 대체로 긍정적이다
- ③ 보통이다
- ④ 대체로 부정적이다
- ⑤ 매우 부정적이다

5. (탄소세) 최근 환경에 대한 관심의 증가와 함께 탄소세의 도입에 대한 논의가 활발해지고 있습니다. 귀하는 현재 또는 가까운 미래에 정부가 탄소세를 도입하는 것에 대해 어떠한 입장이십니까?

- ① 매우 긍정적이다
- ② 대체로 긍정적이다
- ③ 보통이다
- ④ 대체로 부정적이다
- ⑤ 매우 부정적이다

6. 가상자산 양도소득세가 도입된다면 귀하에게 실제 많은 세금이 부과될 것으로 예상하십니까?

- ① 매우 그렇다
- ② 대체로 그렇다
- ③ 보통이다
- ④ 별로 그렇지 않다
- ⑤ 전혀 그렇지 않다

7. 귀하는 정부가 2023년부터 과세할 예정인 가상자산 양도소득세의 도입 과정이 민주적인 절차를 거쳐서 이루어졌다고 생각하십니까?

- ① 매우 그렇다
- ② 대체로 그렇다

- ③ 보통이다
- ④ 별로 그렇지 않다
- ⑤ 전혀 그렇지 않다

8. 귀하가 납부한 세금과 비교하여 정부로부터 받은 혜택수준이 어떻다고 생각하십니까?

- ① 매우 높은 수준이다
- ② 대체로 높은 수준이다
- ③ 적당한 수준이다
- ④ 대체로 낮은 수준이다
- ⑤ 매우 낮은 수준이다

9. 귀하는 적발될 가능성이 전혀 없다면, 세금납부를 회피할 의향이 있으십니까?

- ① 매우 그렇다
- ② 대체로 그렇다
- ③ 보통이다
- ④ 별로 그렇지 않다
- ⑤ 전혀 그렇지 않다

10. 가상자산 양도소득세가 도입되면 경제적 능력이 높은 사람들이 더 많은 세금을 부담하게 된다고 생각하십니까?

- ① 매우 그렇다
- ② 대체로 그렇다
- ③ 보통이다
- ④ 별로 그렇지 않다
- ⑤ 전혀 그렇지 않다

11. 귀하 가구의 한 달 가구소득은 어느 정도입니까?

- ① 100만원 미만
- ② 100~199만원

- ③ 200~299만원
- ④ 300~399만원
- ⑤ 400~499만원
- ⑥ 500~599만원
- ⑦ 600~699만원
- ⑧ 700만원 이상

12. 귀하의 주택 보유 수는 어떻게 됩니까?

- ① 주택을 소유하지 않음
- ② 1주택
- ③ 2주택
- ④ 3주택 이상

13. 귀하의 교육수준은 어떻게 되십니까?

- ① 중졸 이하
- ② 고졸
- ③ 대졸(전문대 포함)
- ④ 대학원 이상

14. 귀하의 이념적 성향을 종합하여 말씀해 주십시오(0은 가장 보수적, 10은 가장 진보적).

보수			중도				진보			
0	1	2	3	4	5	6	7	8	9	10

6~8회차 설문조사(사회 · 인구통계학적 변수에 대한 질문은 위와 공통)

정부는 가상자산에서 발생한 소득에 대해 20%의 세율을 적용하여 과세하는 것을 검토하고 있습니다. 아래 문항을 읽고 답해주세요.

5. 귀하는 가상자산을 어떻게 정의하십니까?

- ① 네이버, 카카오, 페이코 등 새로운 형태의 결제서비스
- ② 로블록스(Roblox)의 로벅스(Robux)와 같이 메타버스 게임에서 사용되는 화폐
- ③ 탈중앙화된 블록체인 상에서 발행되고 거래되는 화폐
- ④ 신용카드와 같은 무형의 지불수단

6. 가상자산에 관한 뉴스를 접한 적이 있습니까? 접한 적이 있다면 뉴스의 내용은 어떤 내용이었습니까?(복수응답)

- ① 가상자산 가격의 높은 변동성
- ② 가상자산 관련 범죄 및 규제정책
- ③ 가상자산 과세관련 정책
- ④ 가상자산의 유용한 활용법
- ⑤ 가상자산의 종류 등 일반적인 정보
- ⑥ 관련 뉴스를 접한 적 없음

7. (가상자산 소득 과세) 정부는 가상자산에서 발생한 소득에 대해 20%의 세율을 적용하여 과세하는 것을 검토하고 있습니다. 귀하는 가상자산 소득에 대해서 과세하는 것에 대해 어떠한 입장이십니까?

- ① 매우 긍정적이다
- ② 대체로 긍정적이다
- ③ 대체로 부정적이다
- ④ 매우 부정적이다

- 7-1. (긍정 응답시) 가상자산 소득 과세에 긍정적으로 생각하는 이유와 근거를 2-3 문장(혹은 그 이상)으로 작성해주세요. (가장 성실히 응답해주시는 분 중 추첨을 통해 추가 리워드를 제공합니다)

- 7-2. (부정 응답시) 가상자산 소득 과세에 부정적으로 생각하는 이유와 근거를 2-3 문장(혹은 그 이상)으로 작성해주세요. (가장 성실히 응답해주시는 분 중 추첨을 통해 추가 리워드를 제공합니다)

8. 귀하는 가상자산에 투자하고 계십니까?

- ① 예
- ② 아니요

8-1. (투자시) 가상자산에 투자하는 금액 규모가 어느 정도입니까?

- ① 100만원 미만
- ② 100~499만원
- ③ 500~999만원
- ④ 1,000~4,999만원
- ⑤ 5,000~9,999만원
- ⑥ 1억원 이상

8-2. (미투자시) 향후 가상자산에 투자하실 의향이 있으십니까?

- ① 예
- ② 아니요

9. (투자의향 있을시) 가상자산에 투자한다면 예상하는 투자 규모가 어느 정도입니까?

- ① 100만원 미만
- ② 100~499만원
- ③ 500~999만원
- ④ 1,000~4,999만원
- ⑤ 5,000~9,999만원
- ⑥ 1억원 이상

3. 세종 품사 태그

대분류	태그	설명	대분류	태그	설명
체언	NNG	일반 명사	어말 어미	EP	선어말 어미
	NNP	고유 명사		EF	종결 어미
	NNB	의존 명사		EC	연결 어미
	NR	수사		ETN	명사형 전성 어미
	NP	대명사		ETM	관형형 전성 어미
용언	VV	동사	접두사	XPN	체언 접두사
	VA	형용사	접미사	XSN	명사 파생 접미사
	VX	보조 용언		XSV	동사 파생 접미사
	VCP	긍정 지정사		XSA	형용사 파생 접미사
	VCN	부정 지정사	어근	XR	어근
관형사	MM	관형사	부호	SF	마침표, 물음표, 느낌표
부사	MAG	일반 부사		SP	쉼표, 가운데딴점, 콜론, 빗금
	MAJ	접속 부사		SS	따옴표, 괄호표, 줄표
감탄사	IC	감탄사		SE	줄임표
조사	JKS	주격 조사		SO	붙임표(물결, 숨김, 빠짐)
	JKC	보격 조사		SW	기타기호 (논리수학기호, 화폐기호)
	JKG	관형격 조사			
	JKO	목적격 조사	NF	명사추정범주	
	JKB	부사격 조사	NV	용언추정범주	
	JKV	호격 조사	NA	분석불능범주	
	JKQ	인용격 조사	한글 이외	SL	외국어
	JX	보조사		SH	한자
JC	접속 조사	SN		숫자	

자료: 김수현 외(2020), p. 69

4. 주제별 워드 클라우드

말뭉치(corpus)를 주제별로 묶어 단어의 빈도를 도식화하는 워드 클라우드(word cloud)로 텍스트 자료를 시각화할 수 있다. 워드 클라우드는 빈도를 나타내는 글자의 크기로, 단어를 나타내는 그림으로 문서 내에 자주 등장하는 단어와 대략적인 빈도를 직관적으로 알 수 있다. [부그림 1]~[부그림 6]은 연도별 기사와 댓글을 LDA로 추출한 주제에 따라 분류하여 그린 워드 클라우드이다.

[부그림 1] 토픽별 워드 클라우드: 2018년 기사와 댓글



주: <표 V-2>, <표 V-3>의 키워드를 시각화한 것으로 해당 표에서 분류한 주제와 동일한 수의 워드 클라우드가 수록됨

자료: 저자 작성

[부그림 2] 토픽별 워드 클라우드: 2019년 기사와 댓글

<기사>



<댓글>



주: <표 V-2>, <표 V-3>의 키워드를 시각화한 것으로 해당 표에서 분류한 주제와 동일한 수의 워드 클라우드가 수록됨

자료: 저자 작성

[부그림 4] 토픽별 워드 클라우드: 2021년 기사와 댓글

〈기사〉



〈댓글〉



주: 〈표 V-2〉, 〈표 V-3〉의 키워드를 시각화한 것으로 해당 표에서 분류한 주제와 동일한 수의 워드 클라우드가 수록됨

자료: 저자 작성

[부그림 5] 토픽별 워드 클라우드: 2022년 기사와 댓글

〈기사〉



〈댓글〉



주: <표 V-2>, <표 V-3>의 키워드를 시각화한 것으로 해당 표에서 분류한 주제와 동일한 수의 워드 클라우드가 수록됨

자료: 저자 작성

텍스트 분석을 이용한 조세정책에 대한 인식 연구

배진수 · 박정흠 · 김수현

본 연구에서는 인터넷 미디어 여론이 조세정책에 대한 국민적 선호를 잘 반영하고 있는지를 텍스트 마이닝 방법론을 통해 정성적, 정량적으로 분석해 보았다. 연구의 범위는 인터넷 미디어에서 활발하게 논의된 적이 있는 가상자산 기타소득세의 신설에 관한 여론을 분석하는 것을 중심으로 이루어졌다. 먼저 개방형 주관식 설문조사를 통해서 가상자산 기타소득세 신설에 관한 긍·부정 의견을 텍스트 형태로 수집하였다. 이와 동시에 해당 정책에 대한 네이버 뉴스 기사, 네이버 뉴스 댓글, 7개 방송사의 유튜브 뉴스의 댓글을 수집하여 인터넷 미디어 여론을 수집하였다. 이후 다양한 텍스트 방법론을 통하여 설문문의 텍스트와 인터넷 미디어 텍스트를 비교 분석하였다.

연관규칙학습과 의미망 분석 방법론을 통해 텍스트들의 특징을 정성적으로 분석해 본 결과는 다음과 같다. 가상자산 기타소득세 신설에 관한 설문 응답의 경우 사용되는 어휘가 정제되어 있었으며, 해당 정책에 대한 긍·부정의 선호에 대한 의견을 짐작할 수 있는 단어들을 많이 발견할 수 있었다. 네이버 뉴스의 경우 설문과 같이 사용되는 단어들은 정제되어 있으나, 정책에 대한 긍·부정의 의견을 얻을 수 있을 만한 단어를 발견하기는 어려웠다. 한편 네이버 뉴스 댓글이나, 유튜브 댓글의 경우는 사용되는 단어들이 산발적이고 조세정책과 무관한 어휘들이 자주 등장하는 등 잡음이 많은 것

을 확인할 수 있었다. 하지만 댓글에서 사용된 단어들은 해당 조세정책에 대한 긍·부정의 견해를 짐작할 만한 어휘들을 일부 포함하고 있어 작성자의 의견을 어느 정도 유추할 수 있을 것으로 판단되었다.

설문과 인터넷 미디어 텍스트에 대한 정량적인 분석은 인터넷 미디어 여론이 조세정책에 대한 실제 선호에 비해서 심하게 편향되어 있을 가능성을 시사하였다. 잠재 디리클레 할당(Latent Dirichlet allocation, LDA)을 이용한 토픽 모델링을 통해 텍스트들이 구성하고 있는 주제들의 구성비를 파악해 본 결과, 인터넷 미디어 여론들은 분석대상이 되는 가상자산 기타소득세에 관한 주제 이외에도 부동산 관련 주제, 정부 정책에 대한 불만, 금융투자소득세, 정치인 등에 대한 주제를 상당히 많이 포함하고 있었다. 반면 설문문의 경우에는 해당 정책에 대한 긍·부정의 견해와 밀접한 관련이 있는 주제를 형성하고 있는 것을 확인하였다. 한편 가상자산 기타소득세에 긍·부정의 견해가 명확한 설문문의 텍스트를 트랜스포머 모형으로 학습하여 감성분류기를 만들어 인터넷 미디어 여론을 감성분석하는 경우, 긍정 비율이 약 16%, 부정 비율이 약 80%로 나와 긍정 비율이 69.3%인 설문조사에 비해 심한 편향성을 나타내는 것을 확인하였다. 이에 대한 추가적인 분석은, 이러한 편향성이 발생한 이유에 대해서, 해당 정책에 대한 인터넷 여론을 조성하는 이들 다수가 조세정책에 부정적인 의견을 가진 사람이었기 때문으로 추론할 수 있는 결과를 제시하였다.

A Text-analysis on the Perception of Taxation: The Case of Virtual Asset tax in Korea

Jinsoo Bae, Cheonghum Park, and Soohyon Kim

The paper investigates whether public opinions on taxation from online media well represent the true preference of taxpayers, using qualitative and quantitative text-analysis methods. To this end, we collected yes-no opinions on the virtual asset tax planned to be levied by the Korean government from a survey along with written rationales for their opinions. We also collected online public opinions on this taxation policy from online media, such as Naver news articles and its user-generated comments, plus comments on news clips on Youtube in the form of text. Then we compared the texts from the survey and online media using several text-analysis methods.

Qualitative analysis using association rule learning and semantic network analysis shows that the texts from the survey and online media are quite different in their characteristics. The texts from the survey are well organized, and the vocabulary used is informative to infer the respondents' preferences on the taxation policy. Texts from news articles are also well organized, but they do not contain words that convey the writers' sentiment on the policy. User-generated comments are not

particularly well organized, but some words used seem to reflect the writers' attitude towards the tax policy after carefully filtering the texts.

Subsequent quantitative analyses of the texts from the survey and online media suggest that the public opinions on taxation policy reflected in online media can be seriously biased. Topic modeling using Latent Dirichlet Allocation(LDA) shows that the texts from online media consist of various topics, including real estate issues, dissatisfaction with the government, financial investment income tax, and politicians, which are not closely related to the planned implementation of virtual asset tax. In contrast, the texts from the survey are composed of topics closely related to rationales for or against virtual asset tax. In addition, a sentiment analysis using a transformer model trained by the survey texts classified the text from online media as supportive 16% of the time and unsupportive 80%, while 69.3% of respondents in the survey are supportive of virtual asset tax. This stark contrast suggests that public opinions on the taxation policy from online media over-represent taxpayers who are against the policy. Further analysis suggests that this bias occurs because more taxpayers who are against the policy leave their opinions online than those who are not.

■ 저자약력

배진수

연세대학교 경제학과 졸업
미국 The Ohio State University 경제학 박사
전, 한국전력공사 경영경제연구원 일반연구원
현, 한국조세재정연구원 부연구위원

박정흠

서울대학교 경제학부 졸업
미국 Boston University 경제학 박사
현, 한국조세재정연구원 부연구위원

김수현

서울대학교 경제학부 졸업
연세대학교 경제학 박사
현, 전남대학교 경제학부 교수

자료 수집 및 정리

윤다솜 한국조세재정연구원 연구원
이아름 한국조세재정연구원 연구원
최예나 한국조세재정연구원 연구원

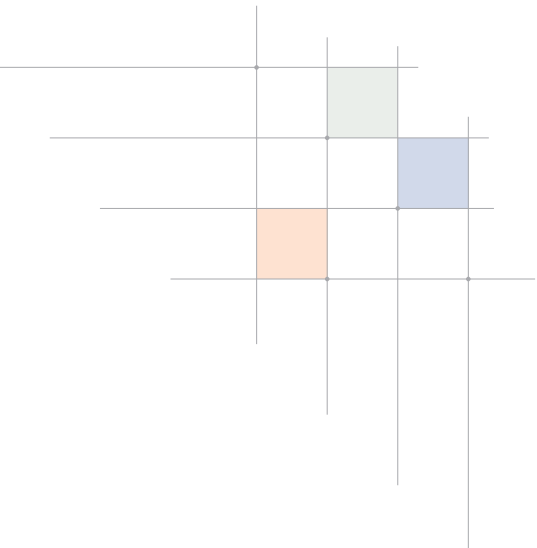
연구보고서 22-07

**텍스트 분석을 이용한 조세정책에
대한 인식 연구**

발행	행	2022년 12월 30일
저자	배진수 · 박정흠 · 김수현	
발행인	김재진	
발행처	한국조세재정연구원	
주소	30147 세종특별자치시 시청대로 336	
전화	(044)414-2114(대)	
홈페이지	www.kjpt.re.kr	
등록	1993. 7. 15. 제2014-24호	
정가	16,000원	
조판 및 인쇄	일지사	
I S B N	979-11-6655-187-1	

© 한국조세재정연구원 2022 * 잘못 만들어진 책은 바꾸어 드립니다.

본 보고서는 친환경 용지를 사용하여 인쇄되었습니다.



KOREA INSTITUTE
OF PUBLIC FINANCE

kipf 한국조세재정연구원

30147 세종특별자치시 시청대로 336

TEL: (044)414-2114(대) www.kipf.re.kr



9 791166 551871
ISBN 979-11-6655-187-1