

과학기술이슈에 대한 일반인의 인식분석: 토픽모델링을 활용한 원자력발전 사례*

최현도** · 안종욱***

<목 차>

- I. 서론
- II. 문헌분석
- III. 빅데이터 분석: 토픽모델링 방법론
- IV. 분석결과
- V. 결론 및 시사점

국문초록 : 과학기술 수용자는 과학기술 변화과정을 결정하는 중요한 이해관계자이다. 하지만 이들이 과학기술이슈에 대해 어떤 생각을 가지고 있는지 확인하는 기존의 방법들은 많은 노력과 시간이 필요한 것으로 알려져 왔다. 본 연구에서는 빅데이터 분석에 널리 사용되는 토픽모델링을 활용해 온라인 토론장에 게시된 글을 분석하여 한국인의 원자력발전에 대한 인식을 알아보고자 한다. 이것은 거시이슈들이 일반 과학기술 수용자 인식구조에 어떤 영향을 미치며, 변화된 인식구조의 지속성을 이해하는데 도움을 줄 수 있다. 빅데이터를 이용해 실시간으로 특정 과학기술이슈에 대한 일반인의 인식을 파악한다면, 과학기술 수용자(일반인)와 공급자(전문가) 집단 사이의 인식간극을 줄이는데 도움을 줄 수 있을 것으로 기대한다.

주제어 : 토픽모델링, 빅데이터, 과학기술이슈, 수용자, 원자력발전

* 이 논문은 2015학년도 조선대학교 학술지원비의 지원을 받아 연구되었으며, 2015년 제17회 경영관련학회 통합학술대회의 기술경영경제학회 분과에서 발표한 내용을 발전시킨 것임.

** 조선대학교 경영학부 조교수 (hyundo_choi@chosun.ac.kr)

*** 한국기술교육대학교 기술혁신경영연구소 연구교수, 교신저자 (jw.ahn@koreatech.ac.kr)

How does the General Public Understand Science and Technology Issues?: A Case on the Nuclear Power Issue Using Topic Modeling Approach

Hyundo Choi · Jongwuk Ahn

Abstract : The general public is a key stakeholder in the science and technology domain. However, traditional approaches require substantial efforts and resources to analyze how does the general public understand science and technology issues. We applied the topic modeling, a form of text clustering, to the texts about the nuclear power which were posted on an online space in order to explore the general public's thoughts on the issue. This study investigates the extent to which macro-level events influence understandings of the general public on the science and technology issues and whether these changes in understandings are sustained over time. It examines the possibility of applying topic modeling in narrowing a perception gap between the general public and the experts through a near-real-time monitoring of the public interests and perceptions about the science and technology issues.

Key Words : Topic modeling, Text mining, Science and technology issues, General public, Nuclear power

I. 서론

최근 미국과학진흥협회 (American Association for the Advancement of Science, AAAS)와 Pew Research Center는 공동으로 일반인과 과학자를 대상으로 높은 관심을 받고 있는 과학기술이슈에 대해 설문조사를 진행했다. 발표된 설문조사의 결과에서 우리는 과학자 집단과 일반인이 해당 과학기술이슈에 대해 서로 다른 인식을 가지고 있음을 확인할 수 있었다(Leshner, 2015). 이를테면 유전자변형식품(Genetically Modified Organism, GMO)에 대해 일반인 중 37%만이 안전하다고 느꼈지만, 과학자의 경우 88%가 안전하다고 생각했다. 유전자변형식품이 과학적 근거를 바탕으로 정말로 안전한지 여부를 판단하는 것도 중요하지만, 두 집단 사이의 차이가 왜 발생하고 있는지 그리고 이 간극을 어떻게 줄일 수 있는지도 중요한 주제라고 할 수 있다.

본 연구는 이러한 문제의식에서 시작한다. 현대사회에서 과학기술은 눈부시게 발전하고 있다. 하지만 그 발전 속도만큼 과학기술이슈의 많은 부분이 논쟁적 영역으로 남아있기도 하다. 논쟁은 많은 경우 앞에서 살펴본 예와 같이 전문가와 일반인의 인식차이에서 온다. 두 집단 사이의 차이를 이해하기 위해서는 과학기술 수용자 입장에 있는 일반인이 과학기술을 어떻게 이해하고 있으며, 그것은 어떤 맥락에서 비롯된 것인지 확인하는 작업이 필요하다.

일반인 인식은 미국과학진흥협회와 Pew Research Center가 수행한 것과 같이 설문조사를 통해서 이루어지는 경우가 많다. 하지만 설문조사는 상당히 많은 시간과 자원이 소비되는 일이다. 설문 영역에서 발생하는 오류, 이를테면 설문 대상이나 질문문항 구성에서의 문제도 있을 수 있다. 다행히도 최근의 정보통신기술 발전으로 우리는 다양한 온라인 공간에서 빅데이터를 수집하는 것이 가능해졌다. 토론 커뮤니티나 개인 SNS(Social Network Service) 등이 데이터 수집의 대상이 되는 대표적인 온라인 공간이다. 여기서 추출한 데이터를 분석할 경우, 일반인의 생각을 파악하는 작업을 보다 원활하게 적절한 시점에서 수행할 수 있을 것이다. 실제로 학계는 물론 실무영역에서도 최근 빅데이터는 널리 활용되고 있다(Griffiths and Steyvers, 2004; Zhao et al., 2011).

위 방법론은 과학기술이슈에 대해서도 적용이 가능하다. 본 연구에서는 원자력발전 기술에 대한 한국 일반인들의 인식을 인터넷 토론장 게시글을 통해 파악하고자 한다. 수집한 게시글은 대표적인 빅데이터 문헌분석 방법의 하나인 토픽모델링을 통해서 분석할 것이다. 토픽모델링은 주어진 텍스트의 작성자가 염두에 두고 있는 주제가 무엇이고 해

당 주제의 구체적 내용은 무엇인지를 텍스트에 포함된 중심어를 이용해 분석해준다.

토픽모델링을 통해서 우리는 한국의 일반인들이 원자력 발전과 관련해서 가지고 있는 인식이 무엇이고 거시이슈에 어떻게 반응을 하는지 확인할 것이다. 특히 외부에서 주어지는 거시이슈들 — 이를테면 후쿠시마 원자력 발전소 사고와 같은 것들 — 에 일반인들은 어떻게 반응을 보이며, 이것의 지속성에 대해 검증하는 연구를 진행할 것이다. 본 작업은 빅데이터를 통한 토픽모델링이 사람들의 인식변화를 확인하고 향후 전문가와의 인식차이를 줄이는데 사용할 수 있는 가능성을 탐색하는 사례연구가 될 수 있을 것으로 기대한다.

II. 문헌분석

1. 시민(사용자)참여형 과학기술혁신

과학기술변화 동력에 대한 가장 대표적이면서 고전적인 두 가지 견해가 있다. 한쪽에서는 새로운 과학기술지식의 발견이 혁신활동을 주도하는 중요한 동력이라고 주장한다. 이것이 기술주도형(tech-push) 관점이다. 다른 쪽에서는 과학기술의 발전은 사회적 혹은 시장 수요에 의해서 견인되었다고 주장한다. 이것은 수요견인형(demand-pull) 관점이다 (Mowery and Rosenberg, 1979).

기술주도형에 대한 논의는 혁신이론의 출발과 함께 한다. 1960년대 혁신이론 초창기, 과학기술의 발전은 연구실에서 시작된 아이디어가 시장에서 상품화되는 단선형 모형이 주를 이루고 있었다. 이것은 자연스럽게 기술이 사회적 수요에 선행한다는 기술주도형 관점으로 이어졌다(Godin and Lane, 2013). 반대로 수요견인형은 실험실에서의 연구개발 방향설정이 필연적으로 시장과 사회의 수요를 반영한다는 인지 아래에서 시작된다.

기술주도형과 수요견인형 관점은 1960~70년대를 관통하며 혁신이론학계에서 경쟁을 해왔다. 다소 진부하기는 하지만 오랜 이론적 논쟁의 결과는 두 관점이 서로 보완적으로 작동한다는 것이다. 이러한 결론은 1980년대 Kline의 체인링크모형(chain linked model)과 Nelson, Lundvall 등의 혁신체제(innovation systems)이론에 자연스럽게 반영된다 (Kline, 1985; Lundvall, 1992; Nelson, 1993).

지금까지 수요견인형을 채택한 많은 연구들은 수요를 시장수요(market demand)로 한

정하여 협의적으로 해석했다. 하지만 과학기술이 사회에 미치는 영향은 점점 커지고 있고 시장 요소만을 고려할 때 많은 현상들이 설명되지 못하고 있는 상황에서는 수요의 개념이 시장의 울타리를 벗어나 확장될 필요가 있다는 공감대가 형성되고 있다(Geels, 2004; Godin and Lane, 2013). 그 결과 과학기술에 대한 사회구성원들의 이해와 공감대 형성이 중요한 과제로 등장했다.

Joss and Durant(1995)는 사회구성원이 과학기술 발전의 과정에 직접적으로 참여하는 방법으로 열린 토론회를 제안했다. 이영희(2007)는 한국에서 시민사회단체가 기술영향평가에 참여했던 경험을 정리하였다. 앞서서와 같이 국내외에서 과학기술에 대한 사회적 공감대를 높이고자 하는 여러 시도를 살펴보면 그 과정에서 다양한 문제가 발견되고 있다.

많은 경우 문제는 일반인과 전문가 사이의 과학기술에 대한 인식차이에서 만들어지며, 이것은 한국의 원자력 발전 사례에도 동일하게 적용된다(Blendon et al., 1993; 진상현, 2014). 인식차를 줄이기 위한 과정에서 일반인이 과학기술에 대해 어떻게 인식하고 있는지 이해하는 것은 가장 어려운 문제 중 하나이다(Devine-Wright, 2005; Barry, 2014). 일반인들의 인식변화를 이해하기 위해서는 크게 두 가지 질문이 제기될 수 있다. 첫째는 거시이슈와 일반인들의 인식변화 사이의 상관관계이다. 기존 문헌에 따르면, 둘 사이에 상관관계는 상당히 높은 것으로 나타난다. 이것은 환경문제와 원자력발전에 대한 일반인들의 우려수준을 확인한 연구에서도 밝혀졌다. 예를 들어 사람들 인식에 영향을 미치는 거시이슈로 환경운동과 반세계화운동은 환경문제에 대한 우려수준을 높이는 방향으로 경제불황은 환경문제에 대한 우려수준을 낮추는 방향으로 작동한다(Melis et al., 2014; Neresini and Lorenzet, 2014). 한국의 경우도 후쿠시마 원전사고 이후에 일반인들의 걱정수준이 높아졌음을 확인할 수 있었다(진상현, 2014). 다만, 대부분의 설문조사가 원자력 발전의 위험에 대한 일반인 인식에 초점을 두고 있고, 설문조사가 가지는 한계로 다양한 거시이슈에 대한 일반인들의 즉각적 반응유무를 파악하는 것에는 한계가 있었다.

두 번째 질문은 인식변화의 지속수준이다. 특정 거시환경변화 이후 시간이 지나면서 기존에 사람들이 가지던 인식방향으로 회귀하는 것을 안전성이라고 하는데, 체르노빌 원자력 발전사고 이후 원자력 발전의 위험에 대한 정도가 지속성을 가지지 못하고 안정된 수준으로 회귀한다는 것이 Lindell and Perry(1990)를 통해서 밝혀지기도 했다. 하지만 동시에 거시이슈가 영구적으로 사람들 인식에 영향을 미친다는 연구도 존재한다(Renn, 1990). 이러한 상황에서 양극화 가설을 검토할 필요가 있다. 개인의 사회문화적 배경 때문에 그들이 거시이슈를 받아들이는 정도에 차이가 발생하면서 양극화가 발생하고 있으며(Melis et al., 2014), 후쿠시마 원자력 발전소 사고를 사례로 한국의 일반인들이 가지

는 원자력 발전 수용성 태도에 대해 연구한 논문에서도 양극화가 발견된다(왕재선·김서용, 2013). 하지만 앞에서 지적한 것과 동일하게, 기존 연구들은 원자력 발전의 위험에 대한 인식방향에 관심을 두고 있기 때문에 양극화 이유를 탐구하는데 어느 정도 한계가 있을 수 있다.

2. 담론분석과 빅데이터

담론분석은 글이나 말을 통해 특정 영역에서의 가치나 동향을 관찰하는 방법을 의미한다(Barker, 2000). 개개인은 하나의 사회주체에 대해 자신만의 견해를 가진다. 그런데 인간은 사회적 동물로 자신의 견해를 내부에 담아두기 보다는 타자와 의사소통하고자 하는 욕구가 크다. 이를 해결하기 위해 사람은 자연스레 글과 말을 이용하게 되며, 그 과정에서 필연적으로 개인의 의식이 묻어나게 된다. 즉, 개인이 타인에게 말을 하거나 글을 공유하는 것은 담론을 형성하는 행위이다.

한국에서도 많은 사회과학연구에서 담론분석 방법론이 사용되고 있다(강진숙, 2006; 권선필, 2008; 최현도, 2014). 광의에서 담론분석은 특정 주체가 하나의 담론을 어떻게 규정하고 사용하는가를 이해하기 위한 연구작업을 의미한다. 여기서 담론은 해당 주체가 가지는 사회적 구조와 위치를 반영한다(강진숙, 2006). 하지만 본 연구에서는 일반인들이 특정 과학기술이슈에 대해서 무슨 생각을 하고 있는지 의식동향을 파악하기 위한 협의의 개념에서 담론분석을 이용할 것이다.

담론분석은 크게 두 가지 방법으로 나뉘며, 그 첫째는 양적담론분석이다. 양적분석은 담론분석의 대상을 규칙에 따라 범주화하고 통계적 방법을 사용해 범주 간의 관계를 분석하는 것이다. 이는 연구자의 주관성을 최소화한다는 장점을 가지고 있다(Riffe et al., 1998). 두 번째 담론분석 방법론은 질적담론분석이다. 이는 담론분석 대상의 내용을 분석하여 글에서 주체들이 표현하는 담론의 실질적 맥락(context)을 이해하는 방법이다. 이는 양적분석과 달리 분석대상에 포함된 요소들을 모두 수량화 시킬 수 없다는 가정에서 출발하고 있다(최현도, 2014).

본 연구에서는 양적담론분석 방법 중 하나인 토픽모델링을 채택하여 온라인 토론장의 게시글을 분석할 것이다. 온라인 게시글을 이용한 토픽모델링은 빅데이터 분석의 일종이다. 과학기술 이슈 뿐 아니라 다양한 분야에서 일반인들의 반응과 관심의 변화를 탐색하기 위해서 빅데이터 분석을 활용하기 시작했다. 빅데이터 분석은 기존의 사회과학 분석

방법과는 다른 장점을 가지고 있기에 여러 분야에서 기존 방법론의 보완적 대안으로 여겨지고 있다.

일반인의 생각을 알고자 할 때, 기존의 사회과학 방법론은 대부분 설문조사에 의존했다(박희제, 2005). 그러나 설문 문항에 대한 자기보고 응답은 신뢰하기 어렵고, 따라서 연구의 결과 역시 부정확하게 된다(Neuman et al., 2014). 리커트척도나 응답편향 등에 대한 여러 연구들은 설문조사에서 정확한 응답을 얻기가 매우 어렵다는 것을 강조하고 있다(Schuluf et al., 2008). 특히, 많은 데이터가 축적되면서 설문에서의 오답에 대한 원인들이 분석되고 있다(Ansolabehere and Hersh, 2012). 보건분야에서는 진료 기록 등의 정보가 쌓이면서 비구조화 형태의 텍스트 데이터를 처리하기 위해서 기계학습 기법을 활용한 계산기법이 유용해지고, 기존 통계 분석을 수행하는 것의 중요성은 감소했다(Murdoch and Detsky, 2013). 혁신연구 부문에서도 투입자원을 최소화하여 비용을 낮추면서 기술로드맵과 같이 기업 혁신전략 수립에 필요한 산출물을 도출하는데 빅데이터 분석이 활용되고 있다(최성철 외, 2012). 이렇게 빅데이터 분석이 데이터 수집과 분석의 일반적인 형태가 되어갈 것으로 보이지만, 새로운 방법론이 과거의 방법론을 대체하는 것이 아니라 보완하고 확장하는 것임을 잊지 말아야 한다(Boyd and Crawford, 2012).

Ⅲ. 빅데이터 분석: 토픽모델링 방법론

1. 토픽모델링

토픽모델링은 정형화되지 않는 대규모 문헌자료에서 반복적으로 언급되고 있는 이야기 주제들을 찾기 위해 사용되는 빅데이터 분석 방법론의 하나다. 일반적인 작문의 과정에서 작성자가 특정 주제를 염두에 두고 글을 구성하기 때문에 자연스럽게 관련 중심어를 반복적으로 사용하게 되는 상황을 기본적으로 가정한다. 이런 가정 아래에서 동시에 반복적으로 등장하는 다수의 중심어를 클러스터링 하여 주제그룹을 추정한다. 즉, 특정한 문헌 전체에 서로 관련 있는 단어들의 뭉치인 잠재적 토픽을 추출해준다(Blei et al., 2003).

본 연구에서는 토픽모델링 기법 중에서도 표준으로 자리잡은 LDA(Latent Dirichlet Allocation) 알고리즘을 사용할 것이다. LDA 알고리즘의 기본적 논리구조는 개별 텍스트가 주제그룹에 대한 확률분포를 가지고 있고, 이 분포에 따라 작성자가 적절한 단어를

선택해 나가면서 글을 완성한다고 가정하는 것이다. 여기서의 확률분포를 계산하기 위해서는 기계학습(machine learning)에서 널리 사용되는 베이저안(Bayesian) 확률론을 사용한다.

주제어분포확률에 대한 사전가정을 하지 않는다는 장점 때문에 토픽모델링은 다양한 분야에서 활용되고 있다. Griffiths and Steyvers(2004)는 자연과학의 대표적 학술지인 미국국립과학원회보(PNAS, Proceedings of the National Academy of Sciences)의 초록을 분석하여 시기별로 자연과학계 내에서 주목 받았던 연구주제를 선별하는데 토픽모델링을 활용했고, Zhao et al.(2011)은 트위터와 뉴욕타임즈 기사의 주제비교를 통해 두 매체의 특성을 비교하는데 이용했다. 문헌정보학계에서는 토픽모델링을 여러 용도에 활용하고 있다. Blei(2012)는 ‘Science’와 ‘Yale Law’ 학술지를 대상으로 LDA 기반의 토픽모델을 도출하여, 토픽모델링이 정치학, 심리학 뿐 아니라 계량서지분석에도 활용될 수 있음을 밝혔다. Song and Kim (2013)은 생물정보학 연구분야의 지적구조를 분석했는데, 생물정보학 학계에서 계산적 측면보다 생물학적 측면이 더부각되고 있음을 밝혔다. 또한 동일한 저자가 복수의 이름을 사용하거나 복수의 저자가 동일한 이름을 가진 경우에 발생하는 저자명 모호성 해소를 위해 저자명을 식별할 때도 토픽모델이 사용되었다(Song et al., 2007).

한국에서도 토픽모델링을 적용한 연구들이 증가하고 있으며, 대표적인 것이 박성원 외(2014)가 일반인 1,232명의 한국미래에 대한 설문조사 응답을 토픽모델링으로 분석한 것이다. 응답자의 현재와 미래에 대한 평가를 기준으로 집단을 나누어 토픽모델링을 적용한 결과, 각 집단의 사람들은 한국의 미래에 대해 다른 인식을 가지고 있음을 확인할 수 있었다. 강범일 외 (2013)는 신문 기사를 대상으로 주제 기반 오피니언 마이닝을 시도하였으며, 매체들이 이데올로기에 따라 기사를 보도하는 경향성을 확인하였다. 박자현 · 송민 (2013)은 문헌정보학 분야의 연구동향을 규명하기 위하여 발표 논문 초록을 수집하여 토픽모델링 실험을 수행했으며, 논문 주제 분류, 상승 또는 하강하는 주제, 학술지별 선호 주제 등을 발견했다. 토픽모델링 기법이 적용된 정보 시스템의 제안도 이뤄지고 있다. 정다미 외 (2013)는 사회적 이슈를 다루고 있는 대용량의 뉴스 기사를 수집하여 사회 문제 키워드를 추출하는 시스템의 개발을 제안하였고, 배정환 외 (2013, 2014)는 트위터상의 이슈를 추출하고 시각화하는 트위터 이슈 트래킹 시스템을 설계 및 구축하였다. 과학기술과 관련하여 문헌정보학 분야의 연구자들이 토픽모델링을 다양하게 활용하고 있지만 혁신연구에 적용한 연구는 아직 많지 않다.

2. 분석대상과 자료마이닝

본 연구에서는 과학기술 영역에서 논쟁이 활발한 주제 중 하나인 원자력 발전에 대한 일반인들의 의견을 분석하는 작업을 수행할 것이다. 원자력 발전에 대한 의견은 다양한 통로를 통해 표출될 수 있으며, 그것은 공식적인 미디어부터 개인 SNS에 이르기까지 다양한 형태를 띤다. 이 중에서 본 연구는 인터넷 토론장 게시글에 주목했다. 인터넷 토론장은 누구에게나 열려 있으면서 동시에 개인 SNS 보다는 심도 깊은 논의를 주고받을 수 있다는 장점이 있다.

국내 온라인 토론장 중 방문자수 기준 상위 공간은 ‘다음 아고라’, ‘네이트 판’, ‘디씨인사이드’이다(매경이코노미, 2013.08.24). 본 연구에서는 이 중 다음 아고라 토론게시판¹⁾을 선택했다. 다음 아고라 토론게시판은 다양한 사회이슈에 대한 토론으로 특화된 가장 대표적인 게시판이다. 반면 네이트 판의 경우는 성격이 다음 아고라와 달라 토론게시판이 따로 존재하지 않고 대부분 개인의 일상적 고민을 논의하는 글들로 채워져 있다. 동일한 조건으로 원자력 발전에 관한 글을 네이트 판에서 검색할 경우 그 결과가 아고라의 약 1/8 수준에 그쳤다. 디씨인사이드의 경우는 연령대가 10~30대로 한정되며 토론의 전문성이 타 공간에 비해 부족하다는 점에서 역시 제외하였다.

결론적으로 본 연구는 다음 아고라의 게시글 중 2008년 1월 1일부터 2015년 5월 31일 까지 범위에서 “원자력”과 “발전” 키워드가 동시에 포함된 게시판 글을 웹스크랩 기법을 이용해 2015년 6월 28일에 추출하였다. 최근 10년간 한국 원자력 발전분야에서 주요 사건이라 할 수 있는 UAE(아랍에미리트) 원자력발전시설 수주가 이루어진 2009년을 기준으로 1년이 빠른 2008년을 분석시작년도로 설정했다. 그 결과 총 2,757개의 게시글을 수집할 수 있었다. 이 중 중복해서 올린 글과 북한문제 혹은 원자력 관련 공기업이나 공공기관 민영화 등 원자력 발전과 과학기술적 연관성이 부족한 글 946개를 삭제하여 최종적으로 총 1,811개의 글을 연구대상으로 확정하였다.

3. 데이터 처리 및 분석

게시글 원문은 완결적 문장 형태를 가지고 있다. 하지만 토픽모델링을 위해서는 텍스트가 형태소 단위로 정리되어야 한다. 형태소 단위란 중심어 분석을 위해 일반적인 문장

1) 다음 아고라는 토론게시판 이외에 청원, (일상)이야기 게시판 등으로 구성되어 있다

에서 조사 등을 삭제하고 기본 명사 형태로 정리한 것을 의미한다. 예를 들어, “발전소에서” 라는 어절에서 장소를 지칭하는 조사인“에서”를 삭제하고 “발전소”만을 취하는 것이 형태소 단위라고 할 수 있다. 또한 “신재생에너지”와 같은 복합명사의 경우는 “신재생”과 “에너지”로 분리를 해주는 것 역시 형태소 전환과정에 속한다.

본 연구에서는 한글 형태소 변환기로 국민대학교 강승식 교수의 KLT2000 프로그램을 사용했다(강승식, 2002). KLT2000에서는 한글 형태소 변환을 위해 중요한 두 가지 옵션을 제공한다. 첫째는 불용어(stop word) 사전이다. 형태소 전환 과정에서 불필요한 단어들, 예를 들어 감탄사 등을 삭제하는 기능이다. 두 번째는 “되다”, “하다”와 같이 중요도가 낮은 형태소를 제외하는 옵션이 있다. 본 연구에서는 두 가지 기능을 모두 적용하여 형태소 변환 작업을 수행하였으며, 정리된 문서에서 출처링크는 불필요한 형태소로 간주하여 수작업으로 삭제했다.

앞서의 전처리 과정을 거친 텍스트를 이용해 주제어를 추출하고 그들 사이의 관계를 통해 그룹을 만들기 위해서 LDA 알고리즘을 JAVA로 구현한 Mallet 토픽모델링 패키지를 사용했다 (McCallum, 2002). Mallet을 통해 토픽모델링을 수행하기 위해서는 몇 가지 설정이 필요하다. 우선 샘플링 반복횟수가 그것인데 여기서는 1000번 반복하도록 설정했다. 그리고 주제어 그룹 수 역시 사용자가 임의로 선택을 해야 한다. 사실 이것은 토픽모델링이 가지는 가장 큰 약점이다. 이것을 극복하기 위해 본 논문에서는 토픽그룹을 2개, 3개, 4개로 설정한 분석 결과를 비교하여 가장 적절한 그룹수를 도출하였다. 추출된 토픽그룹은 해당 그룹에 소속된 주제어들을 바탕으로 저자들이 적절한 이름을 부여하였다. 추가적으로 시계열 흐름에 따라 그리고 개별 글에 대한 사람들의 주목도에 따라 주제그룹들의 비율이 어떻게 변화하는지 추적하였다.

IV. 분석결과

토픽모델링은 연구대상이 되는 텍스트에 잠재되어 있는 이야기의 주제를 찾는 기법이다. 사람들은 발언을 하거나 글을 쓸 때 자신이 주장하고 싶은 내용을 마음에 두고 실행에 옮긴다. 토픽모델링은 글에는 직접적으로 드러나지 않지만, 작성자가 분명히 가지고 있는 주제의식을 분석하는데 도움을 준다. 여기서는 원자력 발전에 대해 한국의 일반인들이 어떠한 생각과 의식을 가지고 있는지 알아보기 위해 토픽모델링을 활용하여 인터

넷 토론장 게시글의 주제를 분류하였다.

우리는 본 논문에서 두 가지 사실을 확인하고자 한다. 첫째로 다양한 거시이슈에 대해 일반인들이 어떻게 반응을 하는지 둘째로 변화된 인식구조의 지속성과 양극화 현상을 살펴볼 것이다. 기존연구들이 설문조사를 통해 원자력 발전에 대한 위험에 초점을 두었다면 여기서는 실시간으로 수집할 수 있는 게시판 글을 활용해 위험 뿐 아니라 폭넓은 주제에 대한 사람들의 인식구조를 파악하는데 초점을 두고자 한다.

이것을 위해서는 우선 한국에서 원자력 발전과 관련한 주요 사건들을 바탕으로 사회적 맥락을 살펴보는 작업이 필요하다. <표 1>은 2008년부터 2015년 상반기까지 원자력 발전과 관련된 주요 사건들을 정리한 것이다. 주요사건은 포털사이트 네이버 신문기사 검색페이지에서 ‘원자력’과 ‘발전’을 키워드로 주요일간지 분기별 기사수를 검색한 후, 기사가 급격히 증가한 시기의 주요사건을 추출하는 방식으로 확인했다. 기사를 통해 추출된 주요사건은 다시 복수의 전문가 의견을 토대로 다음과 같이 확정하였다.

우선 2009년 12월에는 한국전력이 UAE 원자력공사와 원전건설 계약을 체결했다(한국전력 홈페이지, <http://home.kepco.co.kr>). 이것은 한국에서 처음으로 수주한 해외 원자력 발전시설 공사로 이명박 대통령이 직접 UAE를 방문해 계약을 체결하면서 주목을 받기도 했다. 2011년 3월에는 일본의 태평양 해역에서 발생한 지진해일에 의해 후쿠시마에 있는 원자력 발전시설이 침수되는 사고가 일어났다. 원자력 발전소의 냉각시스템이 가동을 멈추면서 수소폭발이 일어나 대규모 방사능 물질이 방출된 후쿠시마 핵발전소 사고는 한국 뿐 아니라 전 세계적으로 중요한 사건이라고 할 수 있다(경향신문, 2011.03.12).

2012년 3월에는 한국에서 가장 오래된 원자력 발전시설로 1978년부터 상업운전을 시작한 고리원자력발전소 정전사태가 발생했다. 후쿠시마 사건이 벌어진 이후 한국에서 벌어진 사고였고, 발전소 운영자가 사고 이후의 적절한 보고절차를 따르지 않은 문제가 밝혀져서 큰 주목을 받았다(중앙일보, 2012.03.14.). 2013년 4월에는²⁾ 고리원자력발전소 부품의 납품비리사건이 감사원과 검찰에 의해서 밝혀졌다. 사실 고리원자력발전소는 2007년 설계수명이 다했지만 원자력안전위원회 인가를 받아 연장운전을 하고 있었다. 그렇지만 오래된 원자력 발전소인 만큼 고장이 잦았다. 이러한 고장 원인의 하나로 안전규정을 만족하지 못한 부품들이 공급된 것으로 지목되었고, 그 결과 납품비리가 감사원과 검찰에 의해 밝혀진 것이다(KBS, 2013.04.12). 그리고 가장 최근에는 앞에서 지속적으로 문제가 발생한 고리원자력발전소의 2015년 5-6월 재연장 신청을 앞두고 시민단체들의 폐

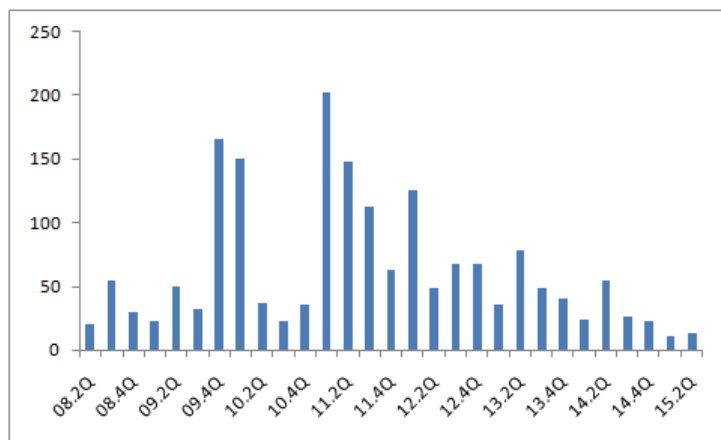
2) 원전부품납품비리 사건은 검찰과 감사원에 의해 2012년부터 수사가 시작됐지만, 확대수사결과가 2013년 4월에 발표되면서 대중매체에 의해 크게 보도되었다.

왜운동이 이슈가 됐다(연합뉴스, 2015.05.21).

<표 1> 한국 원자력 발전과 관련된 주요사건

날 짜	주요사건
2009년 12월	UAE 원자력발전시설 수주
2011년 3월	일본 후쿠시마 원자력발전소 사고
2012년 3월	고리1호기 정전사고
2012년 12월	제18대 대통령선거
2013년 4월	고리원자력발전소 부품납품 비리사건
2015년 5-6월	고리1호기 폐쇄운동

이와 같은 국내외 거시이슈의 발생은 인터넷 토론장에서의 원자력 발전 관련 논의 활성화 정도에도 영향을 미쳤다. <그림 1>은 본 논문의 연구대상으로 확정된 다음아고라 게시글의 숫자를 시기별로 표시한 것이다. 그림을 보면 2009년 4분기, 2011년 1분기에 특히 많은 수의 원자력 발전 관련 글이 올라온 것을 확인할 수 있다. UAE 원자력발전소 수주계약과 후쿠시마 발전소 사고는 그만큼 논쟁거리가 많은 사건이라고 할 수 있겠다. 그리고 이어서 2012년 1분기에 있었던 고리1호기 정전사고로 게시글 숫자가 증가했다. 앞의 세 사건만큼은 아니지만 고리원자력발전소 납품비리사건이 발생한 2013년 2분기에 도 상대적으로는 원자력 발전에 대한 게시글의 빈도수가 높아진 것을 확인할 수 있다. 마지막 고리1호기 폐쇄운동은 주목정도가 낮는데 이것은 전국이슈라기 보다 부산권에 집중된 이슈였고, 폐쇄운동이 5월보다는 데이터 수집범위를 벗어난 6월에 집중되었기 때문으로 해석된다.



<그림 1> 원자력 발전 게시글 빈도

우리는 앞에서 살펴본 사회의 거시적 이슈에 대한 정보를 기초로 토픽모델링 결과를 분석해 볼 수 있다. 토픽모델링 기초 결과는 <표 2>와 같다. 앞에서 설명하였듯이 토픽의 수는 모델링 초기에 수행자가 임의로 지정해야 하는 초기조건이다. 우선 토픽 수를 두 개로 설정했을 때의 결과를 살펴보자. 첫 번째 그룹에는 ‘방사능’, ‘후쿠시마’, ‘안전성’, ‘폐기물’, ‘체르노빌’ 등의 주제어가 속해있다. 이들은 다음아고라에서 개별 게시물들을 분석했을 때 동시에 등장하는 단어들이다. 가까이는 후쿠시마 원전사고, 멀게는 체르노빌 핵발전소 사고를 중심으로 사람들이 논쟁을 벌이고 있음을 알 수 있다. 두 번째 그룹에서 나타나는 중심어는 ‘UAE’, ‘에너지’, ‘연구원’, ‘핵융합’과 같은 것이다. 이를 통해 우리는 한국인들이 원자력 발전이 국가의 에너지 정책 혹은 경제영역에 미치는 영향에 대해 논의하고 있다는 것을 알 수 있다. 여기서 UAE 주제어는 앞에서 살펴본 2009년 12월 UAE의 원자력 발전 설비 수주를 달성한 것이 경제적 논의를 진행하는데 자주 언급되고 있음을 이야기해 준다. 결론적으로 원자력 발전 논의를 두 가지 주제로 구분한 토픽모델링 결과는 사람들이 크게 환경영역과 경제영역으로 나누어 원자력 발전과 관련한 토론을 벌이고 있음을 말해준다.

그룹을 세 개로 정한 경우의 토픽모델링 결과도 함께 추출하였다. 해당 결과 역시 <표 2>에 함께 정리했다. 이것을 보면 원자력 발전의 환경 분야에 대한 논의는 그대로 유지된 반면 경제적 논의가 두 가지로 갈라졌음을 확인할 수 있다. 그것은 원자력 발전 관련 대외경제 이슈와 성장동력 이슈로 이름붙일 수 있다. 대외경제의 경우 반복적으로 등장한 주제어로 ‘UAE’, ‘프랑스’, ‘청와대’, ‘송전탑’ 등이 추출되었고, 성장동력의 경우 ‘자동차’, ‘태양광’, ‘일자리’, ‘이산화탄소’, ‘제조업’, ‘선진국’ 등이 포함되었다.

토픽모델링에서 중요하게 분석할 내용은 특정 주제어가 어디에서 다른 어떤 주제어와 함께 등장하는지에 대한 해석이다. 종종 생각하지 못한 조합의 주제어가 하나의 그룹으로 묶여서 결과가 도출되거나, 혹은 그룹수를 조정했을 때 주제어의 위치가 변경될 경우는 추가 설명이 필요하다. 여기서 주목할 만한 주제어는 핵융합에 관련된 주제어다. 이것은 ‘핵융합’이라는 주제어 자체도 포함되고 ‘KSTAR(초전도핵융합실험장치)’ 같은 핵융합 주요 실험시설이나 ‘ITER(국가핵융합연구소)’ 같은 연구소 이름도 포함된다. 흥미로운 것은 핵융합관련 주제어가 주제그룹을 두 개로 설정했을 때는 환경 주제어와 함께 분류되었지만 주제그룹을 세 개로 설정할 경우는 경제 주제로 구분됐다는 사실이다. 이것은 핵융합이 사람들에게 원자력발전과 관련된 환경문제를 해결할 수 있는 대안기술로 인식되는 동시에 한국 경제성장에 도움이 될 수 있는 성장동력 기술로도 인식되고 있다는 사실을 시사하고 있다. 실제 게시판 글들을 직접 확인해봤을 때도, 핵융합 기술이 두 가지 토픽을 설명하는데 모두 등장하고 있다는 것을 확인할 수 있었다. ‘태양광’이라는

주제어 역시 이와 동일한 방식으로 해석할 수 있다.

전·현직 대통령 이름이 경제분야 토픽에서 집중적으로 등장하고 있다는 사실도 주목할 만하다. 사람들이 환경분야 토픽을 이야기할 때는 대통령의 이름을 자주 언급하지 않는다는 것과 대비된다. 이것은 한국인들이 정치계 혹은 정치인을 평가할 때, 경제이슈가 중요한 기준이라는 인식을 하는 반면 환경문제에 대해서는 이들과의 연관성을 크게 고려하지 않는다고 해석할 수 있다. 원자력 발전이 가지고 있는 위험과 그것에 대한 책임은 경제이슈와 동일하게 정치권에게 있음에도 이 둘을 연관시키지 않는 것이다. 또 하나 흥미로운 점은 ‘박정희’ 전 대통령이 원자력 발전과 대외경제 토픽을 언급하는데 함께 등장한다는 사실이다. 이것은 사람들이 박 전 대통령을 원자력 발전과 관련해 중요한 인물이라고 생각하고 있음을 보여준다³⁾.

마지막으로 토픽 그룹을 네 개로 설정했을 경우는 어떻게 될까? 그 결과는 <표 2>의 마지막에 정리되어 있다. 추출된 주제어를 기초로 각각의 주제그룹을 결정한 결과 앞에서 분석된 세 가지 그룹인 환경, 대외경제, 성장동력 이슈는 동일하게 유지되었다. 여기에 추가된 그룹이 한국정치 토픽이며, 이에 속한 주제어로는 전·현직 대통령 이름과 ‘민주주의’, ‘일자리’, ‘청와대’ 등이 있다. ‘일자리’의 경우는 앞에서는 성장동력 그룹에 속해 있었지만 여기서는 한국정치 그룹으로 분류되었다. 즉, 기술을 통한 일자리 창출이 한국 정치와 정치인을 평가하는데 중요한 기준으로 언급되고 있음을 확인할 수 있다.

<표 2> 토픽모델링 결과

토픽 숫자	주제명	주제어
2	원자력발전의 환경적측면	원자력 에너지 발전소 방사능 원자로 우리나라 후쿠시마 핵발전소 폐기물 우리나라 태양광 핵연료 체르노빌 대한민국 가능성 위원회 신재생 안전성 전문가
	원자력발전의 경제적측면	대통령 원자력 UAE 대한민국 핵융합 이명박 우리나라 연구소 노무현 프랑스 에너지 KSTAR 박정희 발전소 민주당 연구원 ITER 박근혜 프로젝트 사이버
3	원자력발전의 환경적측면	원자력 에너지 발전소 방사능 원자로 우리나라 핵융합 후쿠시 핵발전소 폐기물 우리나라 연구소 핵연료 체르노빌 연구원 KSTAR 위원회 대한민국 ITER 가능성
	원자력발전과 대외경제	대통령 원자력 UAE 대한민국 이명박 발전소 우리나라 프랑스 노무현 박정희 박근혜 민주당 에미리트 민주주의 청와대 올림픽 FTA 러시아 송전탑 공기업
	성장동력	에너지 원자력 발전소 우리나라 자동차 태양광 PWR 시스템 일자리 신재생 대한민국 가능성 선진국 러시아 대기업 이산화탄소 생산량 대규모 이야기 제조업

3) 박정희 정부는 1977년 한국 최초의 상업용 원자력 발전소 고리 1호기를 준공시켰다 (KINGS 홈페이지, <http://www.k-ings.ac.kr>).

	원자력발전의 환경적측면	원자력 발전소 방사능 원자로 에너지 후쿠시마 우리나라 핵발전소 폐기물 우라늄 핵연료 체르노빌 위원회 안전성 한수원 전문가 방사선 핵발전 가능성
	한국정치 (에너지정책)	대통령 대한민국 원자력 이명박 박정희 우리나라 박근혜 연구소 민주당 연구원 민주주의 노무현 청와대 발전소 공기업 FTA 일자리 비즈니스 한나라(당) 위원회
4	성장동력	에너지 핵융합 원자력 우리나라 발전소 KSTAR 대한민국 자동차 ITER 태양광 사이버 연구소 노무현 선진국 신재생 프로젝트 시스템 올림픽 이명박 러시아
	원자력발전과 대외경제	UAE 대통령 원자력 프랑스 PWR 대한민국 이명박 에미리트 러시아 발전소 우리나라 컨소시엄 송전탑 BWR 가능성 외국인 캐나다 기술력 이야기 (웨스팅) 하우스

토픽모델링을 통해 텍스트를 분석하여 그 속에 숨은 주제그룹과 해당 주제어를 추출할 수 있지만, 동시에 개별 텍스트 내 주제그룹의 분포 역시 함께 분석할 수 있다. 지금부터는 시간의 흐름에 따라 사람들이 어떤 주제로 담론을 펼쳐나가는지 분석하여 거시 이슈의 영향력과 그 지속성에 대해 알아보고자 한다. 이를 위해서는 주제그룹의 숫자를 선택해야 한다. 앞에서 우리는 주제그룹을 두 개, 세 개, 네 개로 나누었을 경우의 토픽 모델링 결과에 대해 살펴보았다. 이중 우리는 주제그룹을 세 개로 구분한 경우를 가지고 시계열 분석을 수행했다. 주제그룹을 두 개로 나눈 경우는 분석이 너무 단순해진다는 문제가 있었으며, 네 개 그룹의 경우는 한국정치라는 상당히 일반적인 주제가 독립적인 그룹으로 분류되었다. 그렇지만 주제그룹을 세 가지로 나누었을 때는 세 주제가 원자력 발전과 관련해 서로 독립적이면서도 의미 있는 분석단위가 되리라 판단했다.

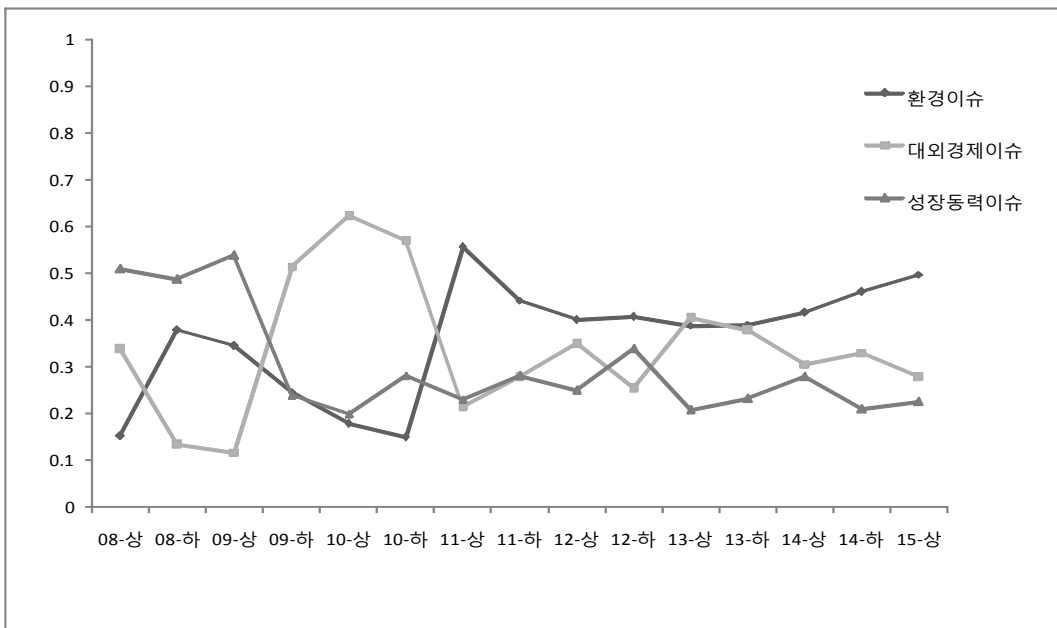
<그림 2>는 개별 텍스트들의 세 개 주제에 대한 시기별 분포 평균값을 도식한 것으로, 시계열 간격은 6개월 단위로 측정했다. 2008년 1월부터 2015년 5월까지를 상반기와 하반기로 나누어 분석했다. 이번 결과를 해석하기 위해서도 앞에서 살펴본 2008년부터 2015년 상반기까지의 원자력 발전 관련 주요 사건 정보가 유용하다. 우선 2008년부터 UAE 원전 수주가 발표되는 2009년 하반기 전까지는 성장동력 주제가 사람들의 인식을 주도한다. 하지만 UAE 원전 수주 이후 상황이 바뀐다. 2009년 하반기부터 약 1년 반 동안은 원자력발전과 대외경제 주제가 사람들 인식을 주도하게 된다. UAE 원전 수주에 성공한 것이 사람들의 원자력 발전에 대한 대외경제적 인식에 상당히 강한 영향력을 발휘했다고 해석할 수 있다.

이러한 흐름은 2011년 상반기에 변화된다. 대외경제 이슈가 하락하고 대신 환경분야 논의가 상승한 것이다. 물론 이것은 2011년 3월 일본 후쿠시마 원전 사고에 따른 것이다. 후쿠시마 사고 이후 원자력발전소의 안전문제 그리고 원자력 발전의 가장 큰 문제 중 하나인 핵폐기물 문제가 사람들의 인식을 환기시켰다고 할 수 있다. 사건 전후의 주제 분

포 비율을 살펴보면 UAE 원전 수주는 대외경제 주제를, 후쿠시마 원전 사고는 환경 주제를 각각 약 40% 이상 상승시켰을 정도로 강한 영향을 미쳤음을 확인할 수 있다.

이후 2013년 상반기에 등장한 고리원전 납품비리의 경우는 UAE 원전 수주를 중심으로 하는 대외경제정책 이슈를 다시 불러일으킨다. 구체적으로 여기에 해당하는 게시글 원문을 찾아보면, 비리가 만연한 원자력발전 산업 때문에 원자력발전소 가동이 정지되거나 UAE 원전 계약이 부정적 영향을 받으면서 대내외경제 측면에서 한국에 경제적 손해가 발생하고 있음을 주장하는 것들이다. 즉, 경제적 측면에서 원전 부품 납품비리를 해석하면서 이를 비판하는 의식흐름이 존재하는 것이다.

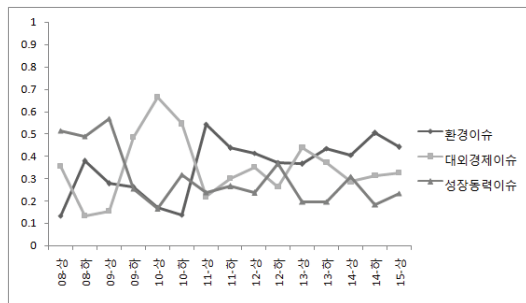
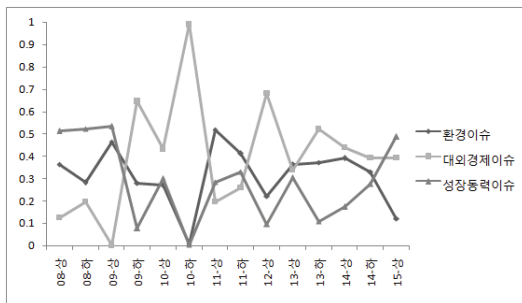
하지만 납품비리 사건은 기존 두 사건에 비해서는 장기적 영향력을 발휘하지 못했다. 약 1년 동안 원자력 발전 관련 대외경제 이슈의 증가를 가지고 왔지만 다시금 환경 논의가 증가하여 후쿠시마 사고 직후의 수준을 회복하게 된다. 다만 주의할 것은 이 결과는 각 주제의 시기별 분포의 비율을 나타낸 것이라는 점이다. <그림 1>에 표시된 절대적 관심도를 함께 본다면, 원전 부품 납품비리 시기의 원자력 발전 관련 주제는 UAE 원전 수주 계약체결이나 일본 후쿠시마 원전 사고 직후 보다는 상당히 낮아져 있기 때문에 원자력발전소에 대한 환경적 관점이 절대적 기준에서 관심을 받고 있다고 해석할 수는 없다.



<그림 2> 3개 주제의 시기별 분포도

추가로 하나 더 고려할 점이 있다. <그림 2>를 통해서 분석한 시계열 분포는 개별 글들을 동일한 가중치를 두고 분석한 것이다. 하지만 게시판에서도 사람들의 의식과 의견을 주도하는 그룹이 있다. 이들을 오피니언리더라고 말할 수 있는데, 오피니언리더 집단의 토픽 분포를 별도로 분석해 보는 것도 새로운 시사점을 줄 수 있을 것이다. 그래서 본 연구에서는 게시판에서 글에 대한 독자들의 찬성횟수를 활용해 오피니언리더의 여부를 구분했다. 다음 아고라에서는 글을 읽은 사람이 글의 내용을 지지할 경우 로그인 이후 찬성의견을 표시할 수 있다. 오피니언리더를 구분하기 위해서는 이 찬성횟수에 대한 기준이 추가적으로 필요하며, 그 최소치를 100으로 설정하였다. 즉 100회 이상의 찬성을 받은 글을 오피니언리더의 글로 보고 별도의 분포를 분석했는데, 그 결과가 <그림 3-1>이다. 오피니언리더의 주요 관심주제를 분명하게 비교하기 위해서 비오피니언리더 집단의 토픽 분포를 <그림 3-2>에 함께 표시하였다. 비오피니언리더 그룹은 찬성을 5개 이하로 받은 글로 정의하였다.

결과를 비교하면 오피니언리더 그룹의 경우 상대적으로 데이터 숫자가 한정적이어서 (총 116개 게시글) 변동성이 비오피니언리더 그룹의 것(총 1271개 게시글)보다 크기는 하지만, 후쿠시마 원전 사고 직후까지의 대체적 패턴은 두 집단이 비슷하게 나타난다. 성장동력에서 UAE 원전 수주 이후 대외경제로 이동했다가, 다시 후쿠시마 사고 이후 환경으로 토픽의 중요도가 이동하는 패턴이 그것이다. 차이는 그 이후에 나타난다. 후쿠시마 사고 이후 1년이 지나면서 오피니언리더 그룹에서는 2008년 초 별다른 거시이슈가 없었을 당시에 주목을 받았던 대외경제와 성장동력 논의가 다시 집중된 반면에 비오피니언리더 그룹에서는 상대적으로 환경에 대한 논의가 높은 분포를 나타내고 있다. 이러한 차이는 2013년 상반기 고리 원전 납품비리 사건을 거치면서 강화된다. 비오피니언리더 집단에서도 고리 원전 납품비리 사건이 발생했을 때 대외경제와 관련된 이슈가 잠깐 부각되었지만 오래가지 못해 약화되었고 대신 환경 분야 이슈가 보다 강화되었다. 두 그룹에서 후쿠시마 원자력발전소 사고의 영향력에 양극화 현상이 벌어진 것이다.



<그림 3-1> 오피니언리더의 3개 주제 시기별 분포 <그림 3-2> 비오피니언리더의 3개 주제 시기별 분포

오피니언리더를 결정하는 찬성 100회라는 기준은 자의적일 수 있다. 연구신뢰도를 높이기 위해서 오피니언리더를 상대적 지표로 정의한 뒤 추가분석을 수행하였다. 찬성 숫자가 높은 상위 10% 게시물(찬성 35개 이상 게시물)에 대해 추가분석을 진행했고, 기존과 유사한 결과를 도출할 수 있었다. 다만, 성장동력과 대외경제 이슈에 집중되는 시점이 찬성 100회 이상이라는 절대치기준으로 오피니언 그룹을 설정했을 때와 비교 시 조금 늦어진다는 점에서는 차이가 있었다. 이를 해석하면 높은 찬성을 받은 게시물일수록 경제적 이슈에 대해 말하는 비율이 높다는 것을 의미한다.

시계열에 따른 세 개 주제그룹 분포 결과를 정리하면 다음과 같다. 첫째, 사람들의 원자력 발전에 대한 의식은 관련 거시사건에 즉각적이고 민감하게 변화한다. 즉, 원자력 발전 관련 거시이슈의 발생에 따라 의식과 관심도 함께 변화하고 토론의 주제가 바로 전환된다. 둘째, 일본 후쿠시마 원전 사고 이후로 원자력발전과 관련된 환경 주제에 대해 사람들의 관심이 높아졌다. 하지만 오피니언리더들은 여전히 경제적 관점에서 원자력 발전을 설명하고 있으며 게시판 사용자들은 이러한 이슈에 높은 관심을 보이고 있다. 보다 적극적으로 설명을 해보자면, 후쿠시마 사고 이후 생태적 감수성이 높고 대규모 기술사고 및 위험에 민감하게 반응하는 사람들은 원자력 발전을 환경이슈와 연관 지어 생각했다. 개인수준에서는 원자력 발전으로부터의 위험을 직접적으로 느낀 사람들이기 때문에 게시판이라는 온라인 공간에서 가장 적극적 참여의 형태인 게시물로써 자신의 인식을 표현했다. 일반 대중의 경우도 후쿠시마 사고에 마찬가지로 영향을 받았겠지만, 직접적 위험을 느끼는 정도가 낮을 경우는 적극성도 낮아진다. 이들은 원자력 발전 문제를 기존과 같이 경제적 관점에서 해석하는 것을 선호하였고, 경제적 관점의 의사를 내포한 오피니언리더의 게시물에 찬성 표시를 하는 정도의 참여에 그쳤다고 해석할 수 있다. 결과적으로 일반인의 원자력 발전에 대한 인식은 후쿠시마 발전소 사고라는 극적인 거시사건 이후에 양극화 된다. 기존연구가 일반인들이 받아들이는 원자력 발전 안전성 정도의 양극화를 살펴본 것과 달리, 본 연구에서는 그들이 원자력 발전을 논의하는데 사용하는 관점과 주제에서 역시 양극화가 발생하고 있음을 확인할 수 있었다.

V. 결론 및 시사점

토픽모델링을 이용해 원자력 발전에 대한 일반인들의 논의 방향을 살펴보았다. 크게 두 가지 결과를 도출했는데, 첫째로 주제어가 어떻게 구분되며 그룹의 수를 증가시키면 주제어 그룹의 구분이 어떻게 변화하는지를 보았다. 두 번째로 이렇게 구분된 주제그룹의 중요도가 시간 흐름에 따라 어떤 변화를 나타내는지 살펴보았다. 이번 분석의 결과는 몇 가지로 요약될 수 있다. 첫째, 일반인들은 원자력 발전을 환경적 관점과 경제적 관점에서 해석하며, 경제적 관점은 다시 성장동력과 대외경제 이슈로 구분될 수 있다. 둘째, 대체로 거시적 사건의 발생에 따라 과학기술 관련 주제에 대한 사람들의 생각과 인식이 크게 변화했다. 이는 2008년 이후 원자력 발전과 관련하여 가장 큰 사회적 사건인 UAE 원전 수주와 일본 후쿠시마 원전 사고에 따른 사람들의 반응을 통해 확인할 수 있었다. 셋째, 오피니언리더들은 과학기술이슈를 환경적 관점보다는 경제적 관점에서 해석하고자 하는 경향을 보이고 있다. 후쿠시마 사고 이후에도 이들은 원자력 발전과 관련해 성장동력이나 대외경제와 밀접한 관련이 있는 담론을 만들었고 이에 대한 일반인들의 선호도가 높았다. 즉, 경제적 관점에 대한 의식흐름의 관성이 크다고 볼 수 있다. 그렇지만 비오피니언리더들에게서는 환경을 중시하는 의식흐름이 분명해지면서 이 둘 사이의 양극화 현상이 발견되기도 했다.

다만 경제적 담론에 대한 선호가 높은 것을 원자력 발전에 대한 일반적 호감으로 해석하는 것은 주의해야 한다. 원자력 발전이 경제적 관점에서 해석되기는 하지만 이것을 긍정적으로 생각하고 있는 것인지 부정적으로 생각하고 있는 것인지는 이번 분석결과에서는 확인하지 못했다. 이는 성장동력 주제에 속하는 주제어에 “태양광”과 “신재생”이 함께 등장했다는 점에서도 확인된다. 개별 게시글 원문을 살펴보면 사람들은 태양광 등 다양한 신재생에너지가 원자력 발전을 대체할 수 있는지 그리고 그것이 보다 한국경제성장에 도움이 될 수 있을지에 대해서 토론을 진행하고 있는 모습을 발견할 수 있었다.

그럼 이와 같은 토픽모델링은 어떤 의미가 있을까? 과학기술이슈를 연구할 때 토픽모델링 방법이 가지는 의의는 다음과 같다. 첫째, 토픽모델링은 사람들의 언행을 통해 그 속에 내포된 의식의 주제를 파악하는데 도움을 주는 도구다. 그렇기에 이것은 일반인들이 과학기술에 대해 어떤 인식을 갖고 있는지 실시간으로 모니터링하여 분석하는데 도움을 준다. 현대 사회의 변화가 가속화되면서 거시이슈 변화가 많아지고, 사람들은 그에 폭발적으로 반응해 토론주제와 인식이 빈번히 전환되고 있는 것이 사실이다. 여기에 더

해 정보통신기술의 발전으로 연구자가 사용할 수 있는 데이터도 폭발적으로 증가하고 있다. 사람들의 인식과정을 판단하기 위해 이들이 인터넷 게시판이나 SNS 등의 채널을 통해서 발언하는 것들을 토대로 그들의 의식흐름을 실시간으로 파악하는 것이 가능하게 된 것이다. 빠르게 변화하는 거시 환경 하에서는 이러한 방법이 오히려 별도의 설문조사보다 적합한 분석방법이라고 할 수 있다.

둘째, 일반인의 과학기술에 대한 인식은 이들과 전문가 집단 사이의 과학기술에 대한 견해차를 줄이는데 활용할 수 있다. 과학기술이 사회에 수용되는 과정은 아직도 이해가 어려운 문제이다(Leshner, 2015). 이것은 많은 경우에 있어서 일반인들이 해당 과학기술에 대해 어떻게 인식하는지 판단하기가 어렵기 때문에 발생하는 경우가 많다(Devine-Wright, 2005; Barry, 2014). 예를 들어 우리가 분석한 원자력 발전의 경우 많은 사람들이 핵융합을 위협한 원자력 발전기술의 대안이자 차세대 성장동력으로 인지하고 있었다. 하지만 핵융합 발전은 상용화까지 많은 시간과 자원이 필요한 기술로, 이에 대한 정확한 이해를 위해서는 적합한 지식이 공유될 필요가 있다.

일반인이 특정 과학기술을 어떻게 이해하는지 확인하는 작업이 전문가와 일반인 사이의 인식격차를 줄이는 중요한 단계이기는 하지만, 전체과정에서 보면 일부분에 지나지 않는다. 추가로 전문가의 인식구조를 확인하는 연구가 필요하고, 둘을 비교하여 구체적인 인식격차 감소방안을 제시해야 한다. 추후 연구에서 인식격차를 줄이는 방안을 제시할 때 몇 가지 유의할 것이 있다. 우선은 일반인들이 가지는 고유한 합리성에 주목해야 한다. 전문가는 이론관점에서 기술적 합리성에 천착하는 반면, 일반인은 자신의 경험에 따른 행태적 그리고 문화적 합리성을 가지고 있다(Plough and Krinsky, 1987; Blendon et al., 1993). 일반인이 특정 생각을 가지게 된 행위적 계기가 있음을 이해해야 하는 것이다. 예를 들어, 본 논문의 연구결과를 보면 성장동력 주제그룹에 ‘PWR(가압수형원자로)’이 주제어로 포함되었다. 이것은 일반인들이 PWR 형식을 BWR(비등수형원자로)보다 안전한 차세대 기술이라고 인식하고 있기 때문이다. 이는 사람들이 BWR방식의 원전인 후쿠시마 발전소 사고를 가장 최근에 목격했기 때문으로 해석할 수 있다. 하지만 전문가들은 안전성 관점에서 두 형식의 우열성은 판단하기 힘들며, 상황에 따라 안전성 정도가 변화할 수 있다는 의견을 가지고 있다(주간경향, 2011.04.05). 즉, 후쿠시마 사고가 행위적 계기로 작용하여 일반인들의 인식에 영향을 준 것이다.

또한, 일반인이 느끼는 정보 비대칭성 문제를 해결할 수 있는 대안 역시 고민될 필요가 있다(Hayenhjelm, 2006; 이세민·김영욱, 2012). 본 연구에서 원자력 발전의 경제이슈와 관련해 등장한 주제어로 ‘송전탑’이 있다. 관련된 논의를 살펴보면 정부가 전체 발전

량에서 원자력 발전 비율을 현 수준에서 유지할 것으로 발표했지만, 전력수요 증가를 가정하기 때문에 원자력 발전소가 추가로 증설될 필요가 있고 그에 따라 원자력발전소에서 수요지로의 전력배송을 위해 송전탑 문제가 지속될 것에 대한 염려에 대한 것이다. 여기서 문제는 정부가 수요예측에 사용되는 정보를 정확히 공개되지 않아 일반인이 정부 전력수요 예측값을 받아들이지 못하고 있다는 사실이다. 이것은 일반인과 전문가 사이의 인식차이를 줄이는 실질적 과정에 시사점을 줄 수 있을 것이다.

본 연구는 게시판 글이라는 비정형적 데이터를 활용해서 일반인들의 과학기술주제에 대한 인식을 확인하려는 시도의 초기 논문이라고 할 수 있다. 방법론이 새롭기 때문에 방법론 자체에 대해서도 다양한 시도가 필요할 것이고, 이것을 활용하는데 있어서도 추후에 관련연구가 계속되어야 할 것이다. 예를 들어 서론에서 소개된 유전자변형식품이나 다른 과학기술이슈에 대해 본 연구의 방법론을 적용할 경우, 원자력 발전에 대한 이번 결과와 어떤 차이점 또는 공통점이 발견되는지 확인함으로써 여러 과학기술별 속성을 파악하고 정책적 시사점을 확장하는데 도움이 될 수 있을 것이다.

참고문헌

(1) 국내문헌

- 강범일·송민·조화순 (2013), “토픽 모델링을 이용한 신문 자료의 오피니언 마이닝에 대한 연구”, 『한국문헌정보학회지』, 제47권, 제4호, pp. 315-334.
- 강승식 (2002), 『한국어 형태소 분석과 정보검색』, 서울, 홍릉과학출판사.
- 강진숙 (2006), ““교육위기” 담론의 의미와 주체구성 방식 연구”, 『한국언론정보학보』, 제33권, pp. 7-52.
- 경향신문, 후쿠시마 원전서 폭발...대표적 발암물질 ‘세슘’ 검출, 2011.03.12.
- 권선필 (2008), “컴퓨터 활용 질적자료 분석을 통한 행정-정치-경영 담론 분석”, 『한국행정학보』, 제42권, 제1호, pp. 345-364.
- 매경이코노미, “온라인 커뮤니티 이대로 좋은가|커뮤니티 춘추전국시대-디시·아고라 대신 일베·뽀뿌가 대세”, 2013.08.24.
- 박성원·황윤하·조규진·서지영·송민 (2014), 『한국인의 미래 인식, 사회분위기, 미래 적응력 조사 (2차년도)』 (조사연구 2014-06-03), 과학기술정책연구원, 서울.
- 박자현·송민 (2013), “토픽모델링을 활용한 국내 문헌정보학 연구동향 분석”, 『정보관리학회지』, 제30권, 제1호, pp. 7-32.
- 박희제 (2005), “한국인의 과학기술자에 대한 인식분석: 세대, 성, 전공계열의 영향을 중심으로”, 『기술혁신연구』, 제13권, 제1호, pp. 1-23.
- 배정환·손지은·송민 (2013), “텍스트 마이닝을 이용한 2012년 한국대선 관련 트위터 분석”, 『지능정보연구』, 제19권, 제3호, pp. 141-156.
- 배정환·한남기·송민 (2014), “토픽 모델링을 이용한 트위터 이슈 트래킹 시스템”, 『지능정보연구』, 제20권, 제2호, pp. 109-122.
- 연합뉴스, 부산NGO “서병수시장 고리1호기 폐쇄 공약 지켜달라”, 2015.05.21.
- 왕재선·김서용 (2013), “후쿠시마 원전사고 이후 원자력 수용성 및 인식구조 변화에 대한 탐색적 분석”, 『한국행정학보』, 제47권, 제2호, pp. 395-424.
- 이세민·김영욱 (2012), “과학 커뮤니케이션 담론 변화에 따른 참여자간 상호지향성 연구”, 『한국언론학보』, 제56권, 제1호, pp. 31-57.
- 이영희 (2007), “기술의 사회적 통제와 수용”, 『경제와사회』, 제73호, pp. 246-268.
- 정다미·김재석·김기남·허종욱·온병원·강미정 (2013), “사회문제 해결형 기술수요 발굴을 위한 키워드 추출 시스템 제안”, 『지능정보연구』, 제19권, 제3호, pp. 1-23.
- 주간경향, 한국원전은 일본원전보다 안전한가, 2011.04.05.
- 중앙일보, 후쿠시마보다 좋다 큰소리치던 고리원전 결국, 2012.03.14.
- 진상현 (2014), “원자력 전문가와 일반인의 인식 격차 분석”, 『한국정책과학학회보』, 제18권, 제4호,

pp. 149-173.

- 최성철 · 김홍빈 · 윤장혁 (2012), “기술로드맵핑을 위한 특허정보의 SAO기반 텍스트 마이닝 접근 방법”, 『기술혁신연구』, 제20권, 제1호, pp. 199-234.
- 최현도 (2014), “과학기술혁신정책 이슈와 학술연구 간의 상호관계연구: 2008-2011년 주요 일간지와 학술지에 대한 질적내용분석을 중심으로”, 『기술혁신학회지』, 제17권, 제4호, pp. 763-785.
- 한국전력 홈페이지, <http://home.kepco.co.kr/kepco/KE/I/htmlView/KEIAHP00107.do?menuCd=F N0102020702>
- KBS, 고리원전 또 비리 적발...13명 기소, 2013.04.12.
- KINGS 홈페이지, <http://www.k-ings.ac.kr/web/kor/ings01;jsessionid=F021E8104DC2B6DF945F 815B232222C0>

(2) 국외문헌

- Ansolabehere, S., and Hersh, E. (2012), “Validation: what big data reveal about survey misreporting and the real electorate”, *Political Analysis*, Vol. 20, No. 4, pp. 437-459.
- Barry, S. (2014), “Using social media to discover public values, interests, and perceptions about cattle grazing on park lands”, *Environmental Management*, Vol. 53, No. 2, pp. 454-464.
- Blei, D. M. (2012), “Probabilistic topic models”, *Communications of the ACM*, Vol. 55, No. 4, pp. 77-84.
- Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003), “Latent dirichlet allocation”, *The Journal of Machine Learning Research*, Vol. 3, pp. 993-1022.
- Blendon, R. J., Hyams, T. S. and Benson, J. M. (1993), “Bridging the Gap between Expert and Public Views on Health Care Reform”, *JAMA*, Vol. 269, No. 19, pp. 2573-2578.
- Boyd, D., and Crawford, K. (2012), “Critical questions for big data”, *Information, Communication and Society*, Vol. 15, No. 5, pp. 662-679.
- Geels, F. W. (2004), “From sectoral systems of innovation to socio-technical systems insights about dynamics and change from sociology and institutional theory”, *Research Policy*, Vol. 33, No. 6-7, pp. 897-920.
- Godin, B., & Lane, J. P. (2013), “Pushes and pulls: Hi(s)tory of the demand pull model of innovation”, *Science, Technology & Human Values*, Vol. 38, No. 5, pp. 621-654.
- Griffiths, T. L. and Steyvers, M. (2004), “Finding scientific topics”, *Proceedings of the National Academy of Sciences*, Vol. 101 (suppl 1), pp. 5228-5235.
- Hayenhjelm, M. (2006), “Asymmetries in Risk Communication”, *Risk Management*, Vol. 8, No. 1, pp. 1-15.
- Joss, S. and Durant, J. (1995), *Public participation in science: The role of consensus conferences*

in Europe, London: Trustees of the Science Museum.

- Kline, S. J. (1985), "Innovation is not a linear process", *Research management*, Vol. 28, No. 4, pp. 36-45.
- Leshner, A. I. (2015), "Bridging the opinion gap", *Science*, Vol. 347, No. 6221, p. 459.
- Lindell, M. K. and Perry, R. W. (1990), "Effects of the Chernobyl Accident on Public Perceptions of Nuclear Plant Accident Risks", *Risk Analysis*, Vol. 10, No. 3, pp. 393-399.
- Lundvall, B.-A. (Ed.) (1992), *National systems of innovation: Towards a theory of innovation and interactive learning*, London: Pinter.
- McCallum, A. K. (2002), "MALLETT: A Machine Learning for Language Toolkit", <http://mallet.cs.umass.edu>.
- Melis, G., Elliot, M. and Shryane, N. (2014), "Environmental Concern over Time: Evidence from the Longitudinal Analysis of a British Cohort Study from 1991 to 2008*", *Social Science Quarterly*, Vol. 95, No. 4, pp. 905-919.
- Mowery, D. and Rosenberg, N. (1979), "The influence of market demand upon innovation: A critical review of some recent empirical studies", *Research Policy*, Vol. 8, No. 2, pp. 102-153.
- Murdoch, T. B., and Detsky, A. S. (2013), "The inevitable application of big data to health care", *JAMA*, Vol. 309, No. 13, pp. 1351-1352.
- Neresini, F. and Lorenzet, A. (2014), "Can Media Monitoring Be a Proxy for Public Opinion About Technoscientific Controversies? The Case of the Italian Public Debate on Nuclear Power", *Public Understanding of Science*, Published online before print.
- Neuman, W. R., Guggenheim, L., Jang, S. M., and Bae S. Y. (2014), "The dynamics of public attention: agenda-setting theory meets big data", *Journal of Communication*, Vol. 64, pp. 193-214.
- Plough, A. and Krinsky, S. (1987), "The Emergence of Risk Communication Studies: Social and Political Context", *Science, Technology, & Human Values*, Vol. 12, No. 3/4, pp. 4-10.
- Renn, O. (1990), "Public Responses to the Chernobyl Accident", *Journal of Environmental Psychology*, Vol. 10, No. 2, pp. 151-167.
- Riff, D., Lacy, S., and Fico, F. (1998), *Analyzing media messages: Using quantitative content analysis in research*, New York: Routledge.
- Schuluf, B., Hattie, J., and Dixon, R. (2008), "Factors affecting responses to Likert type questionnaires: introduction of the ImpExp, a new comprehensive model", *Social Psychology of Education*, Vol. 11, No. 1, pp. 59-78.
- Song, Y., Huang, J., Council, I. G., Li, J., and Giles, C. L. (2007), "Efficient topic-based unsupervised

name disambiguation”, *Proceedings of the 7th ACM/IEEE-CS Joint Conference on Digital Libraries*, pp. 342-351.

Song, M., and Kim, S. Y. (2013), “Detecting the knowledge structure of bioinformatics by mining full-text collections”, *Scientometrics*, Vol. 96, pp. 183-201.

Zhao, W., Jiang, J., Weng, J., He, J., Lim, E.-P., Yan, H., et al. (2011), “Comparing twitter and traditional media using topic models”, In P. Clough, C. Foley, C. Gurrin, G. F. Jones, W. Kraaij, H. Lee, & V. Mudoch (Eds.), *Advances in information retrieval*, Vol. 6611, pp. 338-349, Springer Berlin Heidelberg.

□ 투고일: 2015. 09. 02 / 수정일: 2015. 11. 14 / 게재확정일: 2015. 11. 26