

텍스트마이닝을 활용한 사회위험 이슈 도출*

Deduction of Social Risk Issues Using Text Mining

Jin Young Won**, Dae Gon Kim

Safety Research Division, National Disaster Management Institute, 136 Mapo-daero, Mapo-gu, Seoul,
Korea

Abstract

The signification of the study which is about predicting future risk factors and responding to risks comes to the fore as the society has been rapidly changing. The foresight studies are almost conducted through expert panel, but in the case of the study targeting 'disaster' that implies the uncertainty and complexity, it could reach the limit for expert's subjective judgment. In this study, main topic mining and the deduction of social risk issues were conducted using text mining targeting news data. Social risk issues were examined with climate change adaptation policy. So it was considered that the suggested method could supplement the expert's opinion objectively. Most deducted social risk issues were matched to climate change adaptation policy. So, if the suggested methodology is modified and supplemented, the data based foresight studies for disaster will be conducted effectively.

Key words: social risk, text mining, foresight

국문초록

사회가 복잡해지고 빠르게 변화함에 따라 미래의 위험요소를 예측하고 대응하는 연구의 중요성이 대두되고 있다. 현재 미래예측 연구는 주로 전문가 패널을 구성하여 진행되는데, 재난과 같이 시·공간적으로 불확실성과 복잡성을 함축하고 있는 주제의 경우 전문가의 주관적 판단의 한계에 부딪힐 수도 있다. 본 연구에서는 뉴스를 대

* 이 논문은 2013학년도 안전행정부 국립재난안전연구원 연구과제 결과를 참고하였음.

** Tel. +82-02-2078-7786. Fax. +82-02-2078-7789. E-mail. wjy30228@korea.kr

*** Corresponding author. Tel. +82-02-2078-7786. Fax. +82-02-2078-7789. E-mail. kdgon1609@korea.kr

Submission & Publication Process

Received: May. 20, 2014 / Revised: July. 9, 2014 / Accepted: July. 14, 2014

상으로 텍스트마이닝을 하여 주요 토픽을 추출하고 사회위험 이슈를 도출하였으며, 도출된 사회위험 이슈를 국가 기후변화 적응대책과 비교 검토하여 본 논문에서 제시한 방법이 전문가의 의견을 객관적으로 보완할 수 있는지를 검토하였다. 도출된 대부분의 사회위험 이슈는 국가 기후변화 적응대책과 매칭되었으므로, 추후 본 연구에서 제시한 방법론을 수정·보완한다면 데이터 기반의 미래 재난 예측 연구를 효율적으로 수행할 수 있을 것으로 사료된다.

주제어: 사회위험, 텍스트마이닝, 미래예측

1. 서론

1. 연구 배경 및 필요성

현대 사회는 개체간의 연결을 넘어서 집단 및 시스템의 연결, 그리고 그것이 만들어내는 사회와 문화, 경제 등 여러 분야 사이의 연계에 의하여 변화하고 있으며, 다양한 차원에 질적으로 복잡성이 더해지는 추세이다. ‘미래연구’는 이러한 변화를 미리 예측하고 이에 대한 정책적 전략을 수립하여 대비하기 위한 학문으로, 현재까지 국가경쟁력 강화를 위한 목적으로 국가 과학기술의 미래를 전망하거나 기술을 예측하는 분야에 주로 이용되어져 왔다. 새로운 도구 및 조직, 직업, 기술의 등장과 사회 구성요소의 다양화·전문화는 사회 복잡성을 증가시키고 있고, 사회를 구성하는 각 하부 시스템은 서로 다른 속도로 복잡성이 증가한다. 하부 시스템간의 발달 속도가 서로 달라 그 격차가 크게 벌어지거나 시스템 수용 한계를 초과하는 이벤트가 벌어질 경우 사회 위기 또는 붕괴현상이 발생할 수 있으며, 이러한 사회 위기가 심화되어 국민의 생명 및 재산과 국가에 막대한 피해를 주는 경우를 재난으로 정의한다. 미래사회의 변화가 새로운 위기 또는 붕괴를 가져왔을 경우 국민과 국가의 안전에 미치는 파급력이 엄청남에도 불구하고, 전술한 바와 같이 미래연구는 주로 과학기술 발전에 이용되고 있을 뿐 국가위기 및 재난분야에 적용된 사례는 많지 않다. 따라서 본 연구에서는 사회 위기 또는 붕괴를 나타내는 극단적 미래 관점에서 미래예측 연구를 재난분야에 접목하고자 하였다.

미래예측은 미래를 분석하여 위협요소를 최소화하거나 피하고 기회요소를 활용하여 극대화하는 것이라고 할 수 있다. 이를 위해 수집된 데이터를 가정과 전문성을 통해 다양한 예측 기법을 적용하여 핵심 이슈를 파악하고 예측결과를 도출하는 과정을 거쳐야 한다. 여기서 미래 초점이 되는 이슈를 확인하고 핵심적인 요인을 파악·분석·평가하는 것을 환경스캐닝이라 한다. 즉, 환경스캐닝이란 문헌검토와 인터넷 검색, 전문가 패널, 컨퍼런스 모니터링 등 다양한 방법을 통하여 관심 주제와 관련된 키워드를 도출하고 그 연관성 및 영향력을 파악하여 기회와 위협요인을 도출하는 과정이다. 2000년대까지 미래예측을 위한 환경스캐닝은 전문가 패널 기법 중심으로 수행해왔다. 전문가 패널은 전형적으로 12~20명 정도의 개인들이 모여서 이루어지게 되고, 3~18개월 동안 주어진 토픽 분야(나노기술, 바이오기술과 같은 첨단기술과 관련된 산업 및 경제 관련 주제들)에 관한 미래이슈에 대해 심사숙고하게

된다(김성태, 2007). 이 방법은 특정 분야에 대한 심도 깊은 전문가의 지식과 견해가 필요할 경우 매우 유용하게 사용될 수 있으나, 재난과 같이 시공간적으로 불확실성과 복잡성을 함축하고 있는 주제를 분석할 경우에는 전문가의 주관적 판단의 한계에 부딪힐 수도 있다. 이러한 문제점을 해결하기 위하여, 전문가들의 정성적 의견을 보조할 수 있는 정량적·객관적 데이터의 도출이 요구되고 있고, 최근 텍스트 마이닝은 이러한 기존의 사회과학적 방법들의 한계를 보완하기 위한 방안으로 떠오르고 있다. 특히 정성적 방법으로 정확하게 확인하기 힘든 사회현상의 변화나 현재 주요 쟁점이 되고 있는 사회 문제를 빅데이터 분석을 통해 파악할 수 있게 되었다(정다미 외, 2013). 따라서 본 연구에서는 텍스트 마이닝을 통하여 국내사회에 위협이 될 수 있는 사회위험 이슈를 도출하고 실제 정책과의 검토를 통해 그 가능성을 확인해 보았다.

2. 연구 개요

본 연구의 목적은 미래에 발생 가능한 위험을 예측하기 위하여 현재 우리사회에 당면한 문제를 식별함에 있어 전문가 패널 중심의 주관적 한계성을 보완할 수 있는 객관적인 근거를 제시하는 것이다. 따라서 본 연구에서는 사회적 사건·사고·사실을 다루는 대용량의 뉴스를 대상으로 텍스트마이닝을 활용하여 토픽을 추출하고, 토픽을 구성하는 핵심 키워드의 빈도, 문서수, 토픽 가중치 정보와 원문 검토를 통해 주요 토픽을 도출하였다. 또한 도출된 각 토픽들을 주제로 뉴스와 논문, 보고서 등을 검토하여 미래 사회위험 이슈를 제시하였고, 이를 실제 정책과 비교 검토하여 적용성을 확인하였다.

본 연구에서 이슈 도출을 위해 활용한 텍스트마이닝은 구조화된 데이터베이스를 대상으로 하는 데이터 마이닝과는 달리 구조화 되지 않은 대규모의 텍스트 집단으로부터 새로운 지식을 발견하는 과정을 의미한다(정영미, 2005). 다시 말해, 텍스트마이닝은 비구조화된 텍스트로부터 흥미 있고 유용한 패턴을 찾아내는 과정으로 볼 수 있으며, 정보추출, 정보검색, 자연어 처리, 텍스트 요약, 자동분류 등에서 사용되는 기법들을 결합하여 사용한다. 텍스트마이닝의 적용의 핵심은 대량의 텍스트로부터 이전에 알려지지 않은 숨은 지식을 찾아내는 것이다(이영재 외, 2011). 텍스트마이닝 기법은 여러 연구자들에 의해 현재까지 수많은 알고리즘이 발표되고 있고, 본 연구에서는 범용적으로 사용되고 있는 상용 텍스트 분석 소프트웨어인 SAS사의 Text Miner를 이용하여 텍스트 분석을 실시하였다.

3. 사회위험의 정의

독일의 사회학자 울리히 벡은 위험을 재앙(disaster)과 비교하여 “우리를 위협하는 발생가능성 있는 미래의 사건”이라고 한다. 위험은 불확실하지만 일어날 수 있는 가능성이 있으며, 측정할 수 없는 위험에 대한 두려움이다(Beck, 1994; 한국보건사회연구원, 2013 재인용). 울리히 벡(2006)은 이미 발생한 파괴적 결과와 위험의 잠재적 요소는 분명히 구분되며, 잠재적 요소의 의미에서 위험은 본질적으로

미래의 구성요소를 표현한다고 주장한다. 이것은 부분적으로는 현재 계산 가능한 피해를 미래로 연장시키는 데에 기초하고 있으며, 또 부분적으로는 일반적인 자신감의 상실이나 ‘위험 승수’에 기초하고 있다. 즉, 본래 위험은 예측과, 아직 일어나지 않았지만 다가오고 있는 재난과 관련이 있으므로 본 연구에서는 사회 위기 또는 붕괴를 지칭하는 극단적 미래 관점에서 미래예측 연구를 재난분야에 접목시켜 사회위험 요소를 파악해보고자 한다.

위험은 위험을 야기하는 원인에 초점을 맞추어 자연적, 기술적, 사회적 위험의 세 가지 범주로 구분된다(Jones, 1993). 자연적 위험은 인간이나 인위적인 기술과는 무관한 자연현상의 급변, 천재지변, 혹은 신의 행위로 말미암은 재난(Disasters)으로 이해된다. 기술적 위험은 건물이나 교량의 붕괴, 공장폭발, 오염 등 인공적 산물이나 기술시스템의 문제로 인한 사고들(Accidents)로 간주된다. 마지막으로 사회적 위험은 사기, 절도, 방화, 폭력 등 순수하게 인간 행동만으로 인해 비롯되는 사건들(Incidents)과 관련된다(김원제, 2003). 이는 2014년 2월 7일 인적재난을 사회재난으로 통합하기 이전, 재난을 자연재난, 인적재난, 사회적 재난으로 유형을 분류하던 것과 그 개념이 거의 유사하므로, 위험 요소를 재난 관점에서 해석하기 용이하다. 이에 본 연구에서는 Jones(1993)가 제시한 ‘위험’의 정의를 바탕으로 사회위험 요인을 탐색 및 추출하였고, 추가적으로 한국사회의 특성을 반영한 더 세부적인 위험 요인을 분석하기 위하여 소방방재청(2005)에서 제시한 한국사회의 위험 유형 분류를 적용하였다.

II. 연구의 범위 및 방법

1. 선행연구 고찰

수많은 데이터 속에서 가치 있는 Insight를 찾기 위하여, 해외에서는 이미 텍스트 마이닝을 적용하여 미래이슈를 파악하고 장기정책을 수립하고 있다. 특히 싱가포르의 경우 RAHS(Risk Assessment Horizontal Scanning) 프로그램을 운영하여 데이터 기반의 국가안전 위협요소를 평가하고 환경변화를 탐지하여 새로운 기회요인을 발굴하고 있다. 이 프로그램을 통해 싱가포르는 해안영역에서의 안전확보를 위한 의사결정을 지원하기 위하여 워크시그널과 예상 시나리오를 작성한 바 있고, 조류독감이 싱가포르에 들어옴으로써 벌어질 수 있는 미래 시나리오를 연구하고 대응방안을 마련하였다. 국내에서는 한국행정연구원(2013)에서 소셜 네트워크 분석 및 웹 설문조사 등의 실증연구를 병행하여 위험소통에 적합한 이론과 모형을 탐색하고, 이를 정리 분석하여 재난 및 위험 정책에 대한 제언을 하였다. 이영재 외(2011)는 텍스트마이닝(text mining)을 도입하여 인적재난사고 사례에 대한 신뢰도를 측정하는 절차와 방법에 따라 과거의 인적재난사고사례를 기반으로 새로운 인적재난 전조자료의 위험등급을 추론하는데 기초가 되는 신뢰도 값을 정량적으로 측정하였다. 정다미 등(2013)은 기존 사회과학 방법론이 지닌 문제점을 보완하기 위하여 사회적 이슈를 다루고 있는 대용량의 뉴스 기사를 수집하고

통계적 기법을 통하여 사회문제를 나타내는 키워드를 추출하는 시스템의 개발을 제안한 바 있다.

텍스트 마이닝을 미래예측(특히 기술예측)에 활용한 연구는 최근 활발히 진행되고 있다. 한국과학기술기획평가원(2011)에서는 인터넷과 네트워크 기법 및 텍스트마이닝 기법을 활용하여 건설분야를 대상으로 미래예측에 필요한 키워드를 도출하여 우선순위를 분석하였고, 최진호 외(2011)는 네트워크 기반 특허 분석과 키워드 기반 분석의 한계를 극복하기 위하여 두 방법을 혼합한 특허 키워드 네트워크 기반 분석 방법론을 제시하였다. 이태봉 외(2010)는 INSPEC 데이터베이스에 과학기술 텍스트마이닝을 적용하여 미래 무인전투기술에 대한 지식네트워크 구조와 국방 유망기술을 식별하는 과정을 예시함으로써 구축된 과학기술 정보체계를 이용한 미래 유망기술의 식별 방법론을 제시하였다.

사회위험에 대한 국내 연구는 2000년대 이후 서서히 증가하고 있다. 한국보건사회연구원(2013)에서는 세계화 및 전지구화되고 있는 위기에 대해서 국민이 체감하고 있는 우리 사회의 위험정도와 이에 대한 대비 및 관리능력에 대한 인지도를 파악하여 리스크에 대한 대응방안을 모색하였고, 소방방재청(2005)은 한국사회가 눈부신 고도경제성장의 이면에 “다양하고 복합적인 위험의 증가”라는 부정적 과정이 함께 이루어진 전형적인 위험사회라 정의하며, 실질적인 위험방지 및 안전보장정책으로 이어질 수 있게 하기 위한 사회안전지표를 개발하고 이에 대한 국민안전의식조사를 실시하였다. 홍성만(2012)은 성찰적 근대화론과 위험사회론에서 설정한 위험사회에서의 위험문제를 위험으로부터 보호되어야 할 공공 영역의 확보라는 안전 공공성의 관점에서 고찰하기 위하여, 위험사회의 논의를 살펴보고 공공성의 관점에서 위험문제를 탐색 및 유형화하여 불산가스 누출과 아스팔트 방사능 누출사례분석을 수행한 바 있다.

2. 연구 절차

뉴스는 사실과 정보를 전달해주는 빠른 매개체이므로, 뉴스 분석을 통하여 현 사회의 주요 이슈를 빠르게 파악하고 객관적 분석을 수행하는데 매우 유용하다. 뉴스기사를 분석 데이터로 삼기 위해서는 먼저 크롤러(crawler)를 통해 뉴스를 수집하고 텍스트 형태로 저장한다. 전체 뉴스 데이터 중 분석 주제에 대한 뉴스기사를 선별하여 추출한 후, 이를 대상으로 가중치 계산, 토픽 분석 등의 텍스트마이닝을 수행하여 문서군을 대표할 수 있는 주요 토픽들을 추출하고 분석한다. 추출된 토픽들로 인하여 사회위험 관점에서 발생 가능한 이슈들을 텍스트마이닝 결과와 기사 원문, 문헌 검토를 통하여 도출하고, 이를 현재 발표된 정책과 비교·검토하였다.



<그림 1> 연구 절차

3. 분석 주제 선정

매우 광범위한 주제를 다루고 있는 전체 뉴스에서 사회 위험요인으로 다가올 수 있는 이슈를 일일이 선별하는 것은 많은 분석시간과 노력을 요하게 된다. 또한 필터링 없이 전체 뉴스 데이터를 대상으로 확률적 토픽 모델링을 수행할 경우, 추출된 주요 이슈들이 사회위험과 전혀 연관이 없거나 관련성이 매우 적을 가능성이 있다. 세계경제포럼(World Economic Forum, WEF)은 시대에 따라 전 세계가 직면하는 리스크의 변화를 파악하기 위해 경제·환경·지정학·사회·기술 분야의 리스크 조사 결과를 고찰 및 분석하여 매년 「글로벌 리스크 보고서」를 발표하고 있다(WORLD ECONOMY FORUM, 2012). 본 연구에서는 「2012 글로벌 리스크 보고서」에서 제시된 <그림 2>의 50개의 리스크 중 발생 가능성과 파급효과가 높은 핵심 리스크를 <표 1>과 같이 선별하였고, 전문가 면담을 통하여 리스크 간 연관성이 높고 국내 영향력이 가장 큰 ‘기후변화 적응실패’와 ‘수자원 공급위기’를 분석 대상 리스크로 선정하였다.



<그림 2> 「2012 글로벌 리스크 보고서」에서 제시한 50개 리스크

※ 자료: World Economy Forum(2012).

<표 1> 발생 가능성과 파급효과가 높은 글로벌 리스크 목록

범주	리스크	내용	비고
경제	만성적 재정 불균형	지나친 정부 부채 문제 해결의 실패	
	에너지/농산물가격 극변동성	극심한 가격변동은 주요 원자재 조달을 어렵게 하며, 성장을 더디게 하고, 대중시위를 촉발하거나 지정학적 긴장관계를 조성	
	심각한 소득 불균형	최상위 계층과 최하위 계층 사이의 격차 확대	
환경	기후변화 적응 실패	정부와 기업이 기후변화의 영향으로부터 시민을 보호하고 기업들이 적응할 수 있도록 하는 효과적인 방안을 제정하고 시행하는데 실패	주제 선정
	토양/수로 관리 실패	생태계와 연관산업을 황폐화하는 수준의 삼림벌채, 수로변경, 광물채굴 및 기타 자연환경을 변경하는 프로젝트들	
지정학	글로벌 거버넌스 실패	취약하거나 부적절한 국제기관, 국제적 합의 혹은 네트워크가 각국의 이해관계와 충돌하게 될 경우에 글로벌리스크 해결을 위한 협력을 어렵게 함	
	테러	개인 혹은 비국가단체가 대규모 인적/물적 피해를 유발	
사회	식량부족위기	적합한 양과 질의 식량 및 영양분의 불충분한 불확실한 공급	
	수자원 공급 위기	담수량 및 수질 감소가 식량 및 에너지 생산과 같이 자원 집약적 시스템간의 수자원에 대한 경쟁증가와 맞물릴 때	주제 선정
기술	사이버 공격	범죄자, 테러리스트, 정부주도 또는 정부와 연계된 그룹이 자행하는 사이버공격	

※ 자료: World Economy Forum(2012).

4. 텍스트마이닝 방법

텍스트마이닝의 일반적 절차는 정보수집, 정보처리, 정보추출, 정보분석으로 구분되고, 정보추출 단계에서는 통계적 알고리즘이나 수학적 모델을 통해 유용한 정보를 추출한다. 본 연구에서는 수많은 정보추출 알고리즘 중 가장 범용적으로 사용되고 있는 기법을 활용하여 키워드 가중치를 계산하고, 토픽을 추출하였다.

1) 가중치 계산

문서는 용어의 집합으로 이루어져 있다는 개념을 이용하여 문서를 행렬형태의 비구조화된 데이터로 변형시킨다. 즉 n개의 문서 내에 m개의 용어가 존재한다면, m*n의 용어-문서 행렬로 나타낼 수 있다. 용어의 빈도는 중요성을 나타내는 중요한 수단이 될 수 있다는 개념에 기초하여, 문서의 핵심어를 추출하거나 검색 엔진에서 검색 결과의 순위를 결정하기 위하여 가중치를 계산하고, 여러 연구자에 의해 제시된 많은 방법 중 TF-IDF(Term Frequency-Inverse Document Frequency) 방법을 가장 많이 사용하고 있다. Salton(1975)은 한 문서 내에 자주 출현하는 용어는 그 문서를 대표할 수 있다는 명제를 통해 문서 내의 단어 빈도수 TF 산정방식을 제시하였고, Spark(1972)는 여러 문서에서 동시에 출현하는 용어는 범용적 확률이 높다는 가정 하에 역문헌 빈도수 IDF 개념을 발표하였다. TF-IDF에서 가중치는 TF와 IDF를 곱한 값으로, 본 연구에서는 여러 가지 산정식 중 가장 오랫동안 일반적으로

사용된 방법(TF는 로그스케일 빈도, IDF는 역문서 빈도)을 선택하였으며 이는 식 (1)로 표현할 수 있다.

$$w_{i,j} = TF \times IDF = \log(f_{i,j} + 1) \times \left(\log\left(\frac{N}{n_i}\right) + 1 \right) \quad (1)$$

여기서, $w_{i,j}$ 는 TF-IDF 가중치, $f_{i,j}$ 는 문서 내 용어의 총 빈도, N 은 전체 문서의 수, n 은 용어가 포함된 문서의 수를 의미한다.

2) 차원축소

용어-문서 행렬은 차원이 매우 크고 대부분 0의 값을 가지는 희소(sparse) 행렬이다. 각 문서에 포함되는 용어가 서로 다르기 때문에, 용어-문서 행렬에서 문서의 수보다 용어의 수가 매우 크고, 이로 인하여 계산 시간이 증가할 뿐만 아니라 군집분석을 할 경우 문서간의 거리와 밀도 등의 특성을 정의하는데 어려움이 존재한다. 따라서 효율적 계산을 위해 행렬의 차원축소 과정은 필수적이며, 자료가 연속형인 경우 고차원에서 저차원으로 투영시키는 선형 대수적 기법이 주로 사용된다. 본 연구에서 적용한 SAS Text Miner 모형은 선형 대수적 기법의 대표적인 종류인 특이값 분해(Singular Value Decomposition, SVD)를 이용하여 차원축소를 수행한다.

용어-문서 행렬에 특이값 분해를 적용하는 궁극적인 목적은 문서 군집을 위해서 문서들을 본래 정의된 단어 공간보다 훨씬 작은 차원의 공간에 나타내고자 하는 것이다(오석원, 2011). m 개의 용어와 n 개의 문서로 구성된 행렬 D 는 다음 식 (2)와 같이 근사행렬로 표현할 수 있다.

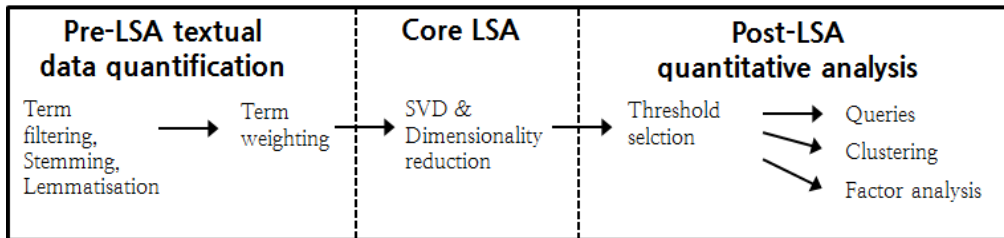
$$\cdot D_k = U \Lambda_k V^T \quad (2)$$

여기서, U 는 $m \times m$ 인 정규직교행렬(orthonormal matrix)이고, Λ_k 는 대각원소가 행렬 D_k 의 특이값이고 대각 원소를 제외한 나머지 원소들이 0인 $m \times n$ 대각행렬(diagonal matrix)이며, V 는 $n \times n$ 인 정규직교행렬이다.

3) 토픽 분석

토픽 모델링은 문서의 주제를 알기 위해 원본 텍스트 내의 단어를 분석하는 통계적인 방법으로, 부가적인 레이블링이나 주석을 달 필요가 없어 대량의 문서 컬렉션을 기계적으로 빠르고 정확하게 처리할 수 있다(정다미 외, 2013). 국내외 여러 연구자들에 의해 발표된 토픽 모델이 존재하며, 주로 사용

되는 토픽 모델에는 LSA(Latent Semantic Analysis), PLSA(Probabilistic Latent Semantic Analysis), LDA(Latent Dirichlet Allocation) 등이 있다. 본 연구에서 적용한 LSA는 본래 정보검색 기술로서 개발되어졌으나, 후에 심리학자들에 의해 용어 구분 및 카테고리 분류 등 인간에 의한 단어 의미를 추출하고 표현하기 위한 방법으로 제안되어졌다(Evangelopoulos, *et. al.*, 2012). LSA를 이용한 토픽분석 방법은 <그림 3>과 같이 정리할 수 있고, 텍스트의 잠재적 의미를 파악하기 위해서는 용어 가중치 산정 및 SVD 계산을 통한 차원축소 과정이 선행되어야 한다. SAS Text Miner는 잠재적 구조를 이해하기 위하여 용어 적재량에 대하여 직각회전 방식을 적용하여 축을 재정의 한다. 절대치가 큰 가중치를 가지는 단어 조합이 존재할 것이고, 이를 제외한 대부분의 단어들은 0에 가까운 가중치를 가지게 되며, 이러한 단어 조합의 문서 한계치와 용어 한계치를 설정하여 토픽을 정의하게 된다.



<그림 3> LSA의 맥락으로 수행되어지는 분석 작업

※ 자료: Evangelopoulos, *et. al.*(2012).

III. 텍스트 분석

1. 정보 수집

본 연구에서는 수많은 뉴스 데이터 중 객관적이면서 영향력이 있는 자료를 선정하기 위하여, 독자적 취재조직으로 수집한 뉴스를 계약을 통해 신문과 방송사, 정부기관 등에 배포하는 연합뉴스¹⁾를 분석 데이터로 선정하였다. 일반적으로 뉴스기사 중 문화·연예 및 스포츠 기사가 많은 부분을 차지하고 있으나, 본 연구에서 다루고자 하는 사회위험과는 연관성이 없으므로 수집 대상에서 제외하였다. 문화·연예·스포츠 기사를 제외한 2012년 5월부터 2013년 10월까지 뉴스 데이터는 총 48만 건으로 집계되었다. 연합뉴스 홈페이지에서는 현재 시점에서부터 1년간의 뉴스자료만을 공개하고 있으므로, 수집 시작일인 2013년 5월의 1년 전인 2012년 5월 이전의 뉴스기사는 분석에 사용하지 못하였다. 각 뉴스기사는 제목과 본문 외에 이미지나 영상, 표 등을 포함하고 있으나, 본 연구의 목적에 적합하지 않은 데이터이기 때문에 제거하였다.

1) <http://www.sisain.kr/news/articleView.html?idxno=17244>(변진경, 2013. 8. 2).

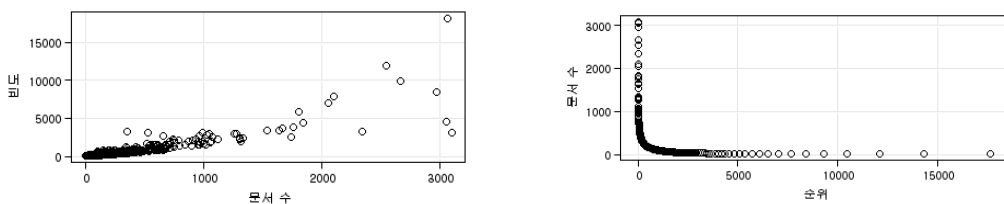
수집된 전체 뉴스기사 중 ‘기후변화 적응실패’ 및 ‘수자원 공급위기’에 해당하는 뉴스 기사를 선별하기 위하여, 관련 문헌 검토 및 브레인스토밍을 통해 각 리스크에 대한 구성요소와 동인, 연관 키워드를 <표 2>와 같이 정리하였고, 이를 통해 분석 대상이 될 뉴스 기사를 추출하였다.

<표 2> ‘기후변화 적응실패’ 및 ‘수자원 공급위기’에 대한 관련 키워드

	기후변화 적응실패	수자원 공급위기
키워드	기후변화, 이상기후, 지구온난화, 해수면 상승, 온실가스 배출, 온실가스 배출량, 온실가스 배출권, 온실가스 감축, 탄소 배출, 탄소 배출량, 이산화탄소 배출량, 기상재해, 기상이변, 저탄소, 녹색성장, 녹색경제, 녹색기술	수돗물, 상수도, 하수도, 중수도, 상수원, 담수량, 저수량, 수질오염, 취수, 녹조, 적조, 가뭄, 물 부족, 지하수, 수자원, 물 수요, 물 수요량, 물 스트레스, 빗물, 물 사용량, 피크 워터, peak water, 물 정점, 물 고갈, 관개용수, 사막화, 해수담수화, 물 파동, 분 전쟁, 물 분쟁, 스마트워터그리드, 강수량, 급수

2. 정보 처리

웹으로 수집한 뉴스 데이터는 비구조화 된 문서이므로, 텍스트 전처리 과정을 통하여 정보 추출이 가능한 형태로 변환하는 작업이 필요하다. 따라서 본 연구에서는 형태소 분석(morphological analysis)을 통해 문서의 입력문자열을 형태소열로 변환하는 작업을 수행하였다. 형태소란 단순어의 어근 및 어미나 조사, 접두사, 접미사 등과 같은 의미의 최소단위로써, 더 이상 분석 불가능한 가장 작은 의미 요소를 의미한다. 전처리된 데이터 중 의미 있는 정보를 추출하기 위하여, 불용어 및 유의어 사전을 구축하여 정보처리에 적용하였다.

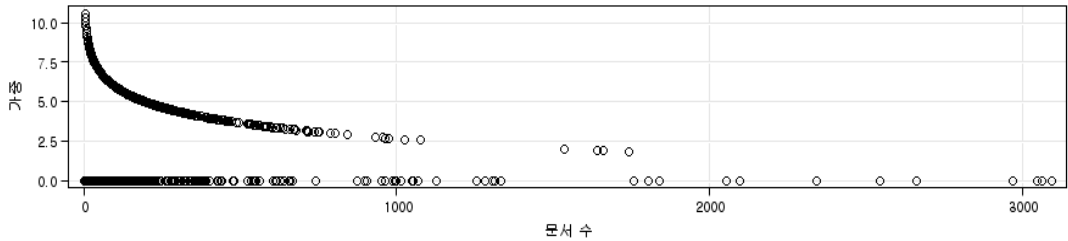


<그림 4> ‘기후변화 적응위기’ 문서군의 용어별 통계결과

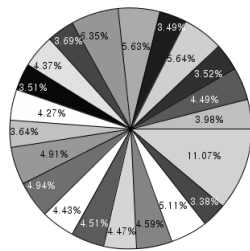
3. 정보 추출

정보 처리 단계를 거치고 나면 각각의 문서는 여러 개의 형태소로 구성된 집합체로 표현되어지고, 전체 데이터는 문서-용어 매트릭스의 DB 형태로 저장된다. 각 용어를 식 (1)에 적용하여 가중치를 계산하고, 용어 간의 연관도를 산정하여 용어의 중요도와 연관성을 분석하였으며, 이후 SVD 계산 및 축

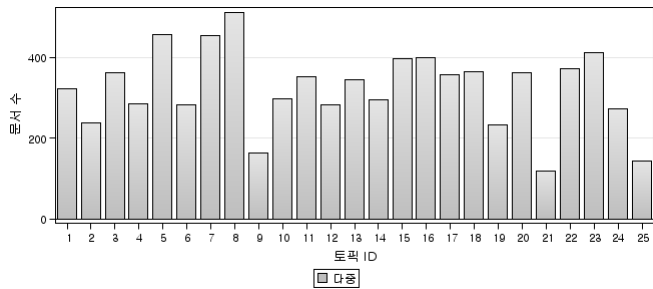
회전을 이용하여 두 리스크 문서군에 대한 토픽을 분석하였다. <그림 5>는 ‘기후변화 적응위기’관련 뉴스기사의 각 용어별 문서수와 가중치를 도시한 결과이고, <그림 6>과 <그림 7>은 토픽 분석 결과이다. <그림 5>에서와 같이 ‘기사’, ‘연합뉴스’, ‘송고’ 등의 분석에 불필요한 용어는 불용어 처리가 되어 출현 빈도에 상관없이 0의 값이 적용되게 되므로, 더욱 효율적인 이슈 분석이 가능하다. 용어들의 조합을 의미하는 토픽은 최소 4개의 문서가 존재할 경우 토픽으로 인정되며, 최대 25개로 분류되어진다.



<그림 5> ‘기후변화 적응위기’ 문서군의 가중치 산정 결과



<그림 6> ‘기후변화 적응위기’ 문서군의 토픽 분류



<그림 7> ‘기후변화 적응위기’ 문서군의 토픽별 문서 수

4. 정보 분석

토픽 분석을 통해 추출된 25개의 토픽이 모두 의미가 있는 것은 아니다. 용어들이 군집을 이루지만 주제와 벗어날 수도 있고, 분석가의 목적에 부합되지 않을 수도 있다. 따라서 추출된 25개 토픽 중 사회위험과 관련된 토픽을 선별하는 과정이 요구되며, 이러한 토픽 선별을 위해서는 ‘위험’에 대한 용어의 명확한 정의가 필요하다. 본 연구에서는 13에서 전술한 바와 같이 여러 연구자의 정의 중 Jones(1993)가 제시한 정의를 기준으로 자연적, 기술적, 사회적 위험 범주로 위험요소를 탐색하였다. 그러나 세 개의 범주에 수많은 위험유형을 포함시키기에는 무리가 있고, 사회의 변화속도와 사회적 특성에 따라 위험환경 또한 변화하기 마련이다. 소방방재청(2005)에서는 현대의 위험사회가 전통시대

와 다른 위협의 요소들을 내장하고 있는데, 그 가장 중요한 요소는 자연과의 관계에서 드러남을 제시하였고, 두 번째로는 인간이 만들어낸 복합적 기술과 문명이 우리가 제어하지 못하는 위험요소가 되어 거대한 구조물과 도시체계, 그리고 산업생산물 속에 내면화되어 생겨나는 위험이라고 정의하였다. 이러한 정의를 기반으로 <표 3>과 같이 한국사회의 위험 유형을 지구적 생태위험, 자연적 재해위험, 국가적 안보위험, 건강의 위험, 경제적 생계위험, 기술적 재난위험, 사회적 해체위험으로 분류하였고, 본 연구에서는 이를 바탕으로 사회적 위험을 다루는 토픽을 선별하여 그 내용을 <표 4>와 같이 정리하였다.

‘기후변화 적응실패’문서군의 주요 토픽은 총 6개로 곡물가격 급등과 홍수·태풍 거대화, 이상기후와 폭염, 전력수급 계획, 작물피해, 생태계 환경변화로 정의할 수 있고, ‘수자원 공급위기’문서군의 주요 토픽은 폭염으로 인한 수자원위기와 4대강과 녹조현상, 수온상승과 적조현상, 상하수도 공급, 가축폐사 피해, 수질오염과 식중독으로 정의할 수 있다.

<표 3> 위협의 유형과 주요 사례

위험의 구조적 수준	위험의 주요 사례
지구적 생태 위험	<ul style="list-style-type: none"> 수질오염(상수원 오염, 공장폐수 등) 대기오염(자동차배기, 황사 등) · 토양오염(화학물질, 쓰레기 등) 해양오염(폐유방류, 적조 등)
자연적 재해 위험	<ul style="list-style-type: none"> 태풍, 폭설, 가뭄, 기근, 폭염, 지진, 쓰나미 등
국가적 안보 위험	<ul style="list-style-type: none"> 한반도에서의 전쟁, 북한 핵문제 중국과 일본의 군사대국화 불특정 다수를 대상으로 한 테러 국가기반체계의 마비(에너지, 통신, 교통, 금융, 의료, 수도 등)
건강의 위험	<ul style="list-style-type: none"> 일반전염병(콜레라, 장티푸스, 결핵 등), 성인병 신종 질병(SARS, 조류독감, 광우병, AIDS 등) 먹거리 안전사고(유해식품, 비위생적 식품, 식중독 등) 생명윤리 침해사고(장기매매, 유전자 조작 등)
경제적 생계 위험	<ul style="list-style-type: none"> 빈부격차 불안정 고용과 취업난 급격한 경기변동 농산물 수입개방
기술적 재난 위험	<ul style="list-style-type: none"> 산업재해, 직업병 자동차교통사고(총돌, 음주운전, 뺑소니 등), 대형교통수단 사고(열차전복, 비행기추락, 선박침몰 등) 시설물 고장에 의한 사고(엘리베이터, 에스컬레이터 사고 등) 화재(가정집, 사무실, 공장 등), 산불 건축물 붕괴사고, 대형폭발사고, 가스사고, 감전 등의 전기사고 화생방사고(화학물질, 방사성물질, 핵폐기물 유출사고) X-ray, 방사선치료의 부작용 컴퓨터바이러스 확산 몰카, CCTV, 도청 등에 의한 사생활 침해 사이버범죄(인터넷을 통한 개인정보유출, 사기, 해킹 등)
사회적 해체 위험	<ul style="list-style-type: none"> 학교폭력, 부정부패, 여아갈등, 노사갈등 살인, 강간, 성폭력, 유괴, 납치, 강도, 절도, 마약 및 알콜 중독

※ 자료: 소방방재청(2005).

<표 4> '기후변화 적응실패' 및 '수자원 공급위기'에 대한 주요 토픽

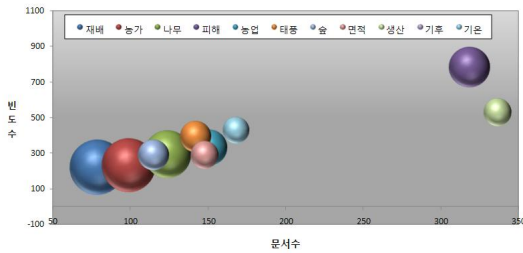
주제	Topic		주요 키워드	주요 토픽	문서 수
	카테고리				
기후변화 적응실패	곡물가격 급등		가격, 곡물, 물가, 시장, 오르다, 상승, 수입, 수출, 옥수수, 증가하다, 기업	주요 곡물생산지역 곡물생산량 감소 및 곡물가격 상승	
	홍수·태풍 거대화		피해, 태풍, 물, 시설, 재해, 침수, 기상, 홍수, 대책, 주민, 관리, 예방, 재난, 집중호우, 하천	자연재해로 인한 피해와 그에 따른 예방·복구비용 증가	
	이상기후와 폭염		폭염, 피해, 환자, 기온, 보험, 기상청, 기상, 질환, 사망자, 태풍	유아, 노년층 등 인명피해와 가뭄·전염병 등 2차적 피해	
	전력수급 계획		전력, 수급, 공청회, 발전, 지경부, 발전소, 전력수급계획, 수요, 화력발전, 에너지, 설비, 건설, 단상, 산업	전력 수요 증가에 따른 제6차 전력수급계획	
	기후변화로 인한 작물피해		재배, 농가, 나무, 피해, 농업, 태풍, 숲, 면적, 생산, 기후, 기온	농가 및 산림피해(2013년 9월 경북 일대 소나무 재선충병)	
	생태계 환경변화		산림, 멸종, 생태, 생태계, 숲, 생물, 나무, 서식하다, 습지, 위기, 식물, 환경, 서식지	북방으로의 수목생장한계선 확장	
수자원 공급위기	폭염으로 인한 수자원 위기		폭염, 열대야, 주의보, 폭염주의보, 더위, 무더위, 경보, 폭염특보, 폭염경보, 특보, 기상청	조류 확산으로 인한 양식 폐사	
	4대강과 녹조현상		조류, 녹조, 남조류, 농도, 수질, 세포, 한강, 클로로필, 낙동강, 물질, 보, 수돗물, 주의보, 정수, 검출되다, 독성, 조류주의보	낙동강과 한강의 녹조현상과 수질오염	
	수온상승과 적조현상		적조, 해역, 양식장, 전남, 생물, 적조주의보, 주의보, 여수, 양식, 황도, 수온, 수산, 어류, 피해, 남해안, 국립수산물과학원, 어민, 폐사하다, 가두리	한반도 연안 일대로 적조 확산	
	상하수도 공급		요금, 상수도, 인상, 수돗물, 하수도, 인상하다, 원가, 시설, 공공요금, 사업, 계량기, 수도, 수도요금, 상수도사업본부	광역상수도 요금인상	
	가축폐사 피해		농가, 닭, 가축, 피해, 폐사하다, 폭염, 축산, 오리, 돼지, 축사, 더위, 재해보험, 불별더위, 사육, 보상, 죽다	양계장 집단폐사	
	수질오염과 식중독		지하수, 사고, 조사, 오염, 수질, 물질, 검출되다, 식중독, 수돗물, 검사, 기준치, 공장, 조사하다, 수질검사	지하수 오염과 조류 독성물질에 의한 수질오염, 식중독	

IV. 사회위험 이슈 분석

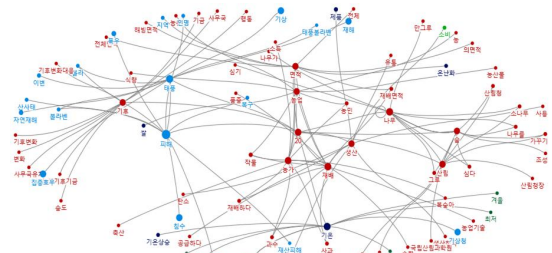
1. 사회위험 이슈 도출

단순히 키워드의 출현 빈도에 따라 의미를 부여하는 것은 한계가 있다. 각 토픽을 구성하는 키워드 중 어떤 키워드가 토픽을 가장 잘 나타내는가를 의미하는 토픽 가중치와 키워드 간 연관도는 토픽의 사회적 의미를 정의하는 데 중요한 역할을 한다. 따라서 본 연구에서는 각 토픽을 구성하는 키워드와 가중치, 연관관계 분석 및 그에 대한 문헌검토를 통해 12개 토픽에 대한 사회위험 이슈를 도출하였다.

<그림 8>은 ‘기후변화로 인한 작물피해’ 토픽에 대한 주요 키워드 분포도이고 <그림 9>는 키워드 간 연관도를 나타내는 키워드 맵이다. 토픽 가중치가 큰 키워드는 ‘재배’, ‘농가’, ‘나무’, ‘피해’, ‘농업’, ‘태풍’, ‘숲’ 순으로 나타났고, 이러한 키워드 분포와 키워드 맵, 기사 원문 검토를 통하여 현재 국내에 기후변화로 인한 농가 및 산림 피해에 대한 우려가 커지고 있음을 알 수 있다. 토픽을 구성하는 다수의 이슈들 중에서 특히 2013년 9월 경북 포항과 경주 일대를 중심으로 소나무 재선충병이 발생하여 산림청에서 긴급방제를 실시한 내용이 가장 중심적인 이슈라 할 수 있다. 이 이슈는 언론에 크게 보도되지는 않았지만 그 피해는 상당했다. 2013년 기후 온난화에 의한 생리적 현상 등으로 소나무 고사목 발생이 급증하여 재선충병이 확산되었고, 산림청은 2013년 9월 기준으로 92만 4,000그루의 고사목이 발생하였다고 발표하였다. 기후변화로 인하여 단기적으로는 농업생산 감소, 재해로 인한 손실, 병해충 발생이 증가할 수 있고, 장기적으로는 기반이 파괴되고 약품 및 비료 사용 증가로 생태계를 악화시켜 생물다양성을 감소시킬 뿐만 아니라 서식지를 파괴하는 등의 문제가 발생할 수 있을 것이다.



<그림 8> ‘기후변화로 인한 작물피해’ topic의 키워드 분포도



<그림 9> ‘기후변화로 인한 작물피해’ topic의 키워드 맵

이와 같이 텍스트마이닝을 통한 키워드 간 중요도와 연관관계 분석은 실제 발생한 이슈뿐만 아니라 미래 사회에 위협이 될 수 있는 이슈를 도출할 수 있다. ‘기후변화로 인한 작물피해’ 토픽과 동일한 방법으로 총 12개 토픽에 대한 사회위험 이슈를 도출하였고, 이를 <표 5>에 정리하였다.

<표 5> '기후변화 적응실패' 및 '수자원 공급위기'에 대한 사회위험 이슈

	카테고리	사회위험 이슈		카테고리	사회위험 이슈
기후 변화 적응 실패	곡물가격 급등 (11)	선물투기의 곡물시장 유입으로 인한 곡물가격 급등 신흥경제국 성장으로 인한 축산물 소비 증가	수자 원 공급 위기	폭염으로 인한 수자원 위기 (17)	수온 상승 및 열대야로 인한 하천오염도 증가
	홍수·태풍 거대화 (12)	이상기온으로 인한 자연재해 거대화 해수면 온도 증가에 따른 해안 도시 태풍 피해 위험도 증가		4대강과 녹조현상 (18)	남조류의 환경 적응력 및 형태적 변이로 인한 인체 유해한 독성물질 생성 중국 거대 녹조 국내 유입
	이상기후와 폭염 (13)	폭염으로 인한 사망자 수 증가		수온상승과 적조현상 (19)	해수온도 상승으로 인한 유해조류 독성 강화 유해조류 이매패류에 축적
	전력수급 계획 (14)	탈핵추세와 저탄소 에너지정책		상하수도 공급 (110)	공공요금 상승에 대한 국민의 불신 가중
	기후변화로 인한 작물피해 (15)	- 농업생산 감소 및 병해충 발생 증가 약품·비료사용 증가로 생물 다양성 감소 및 서식지 파괴		가축폐사 피해 (111)	폐사가축 매몰에 따른 침출수 유출
	생태계 환경변화 (16)	산림 식생대 이동 및 생태계 구조·기능 변화		수질오염과 식중독 (112)	녹조 독성물질로 인한 수돗물 오염 환경지속성 및 스트레스 저항성이 큰 식중독 미생물 증가

2. 국가 기후변화 적응대책 비교 검토

환경부에서는 기후변화의 악영향으로부터 국민의 생명·재산을 보호하고, 안전한 한반도를 만들기 위하여 저탄소녹색성장기본법에 의거하여 2010년 국가 기후변화 적응대책을 수립하였다. 「국가 기후변화 적응대책(2011~2015)」은 건강, 재난·재해 등 10개 부문 대책에 13개 중앙부처가 참여하고 환경부가 총괄한 대책으로, 정부부처 실무협의회와 분야별 전문가 간담회, 심포지엄, 관계부처 협의를 통해 수립되었다. 국가 기후변화 적응대책은 정책 실무자와 분야 전문가들의 지식의 집결체라 할 수 있으므로, 국가 기후변화 적응대책의 분야별 대책과 <표 5>의 사회위험 이슈를 비교하여, 본 논문에서 제시한 방법이 전문가의 의견을 객관적으로 보완할 수 있는지를 검토하였다.

국가 기후변화 적응대책과 본 논문에서 제시한 사회위험 이슈를 비교한 결과, <표 6>과 같이 대부분의 사회위험 이슈가 부분별 기후변화 적응대책과 상응하는 것을 확인하였으나, 간접적인 적응기반 대책인 기후변화 감시예측과 적응산업·에너지, 교육·홍보 및 국제협력 분야의 대책은 사회위험 이슈와 중첩되지 못하는 것으로 나타났다. 기후변화 적응대책과 매칭되지 못한 사회위험 이슈는 '곡물가격 급등'과 '전력수급 계획'로 인한 사회위험 이슈이다. '곡물가격 급등'토픽으로부터 유발될 수 있는 사회위험 이슈는 곡물 주요 생산지의 기상이변과 더불어 선물투기의 과도한 곡물시장 유입과 바이오 연료 사용, 신흥 경제국의 축산물 소비 증가가 곡물의 국제거래가격을 치솟게 하는 것이다. 이러한 곡물가

격 급등현상은 국내의 광범위한 식품 가격의 인상에 영향을 미치게 되고, 가격 인상은 개인소비를 감소시킬 것이며, 기업의 경우 원료 인상분을 소비자 가격에 반영하지 못한다면 수익성이 떨어지게 된다. 한국보건산업진흥원(2012)에서는 한국이 곡물자급률이 낮은 수준이며 국제 곡물가격의 변동성에 취약한 식품산업 구조를 이루고 있음을 배경으로 국제곡물가격 상승에 따른 2013년 국내 식품물가 전망을 발표하였다. 한국보건산업진흥원(2012)에 따르면 최근 미국, 남미, 러시아 등 주요곡물 생산국에서 극심한 가뭄이 발생함에 따라 국제곡물가격이 2007~08년 애그플레이션 때와 유사한 단기간 급등락 현상을 보이는 등 가격변동성이 확대되었다. 미국 농무부(USDA)는 2012년 세계 곡물수급전망 보고서에서 2012/2013년도 세계 곡물 생산량이 3.5%, 소비량 1.2%, 재고량 10.1% 감소할 것이라 발표하였고, 이로 인해 국제곡물가격이 상승될 것으로 전망하였다. 실제로 2012년 하반기에 국내 식품제조업체들은 제품가격을 3~20% 인상하였으며, 이는 2010년 하반기부터 2011년 상반기까지 상승하였던 국제곡물가격이 반영된 결과로 볼 수 있다고 발표하였다. 이와 같이 기후변화로 인해 국제곡물가격이 국내 식품물가에 더욱 큰 영향을 미침에도 불구하고, 국가 기후변화 적응대책의 농업분야에서는 작물 재배기술 개발과 농업용수의 효율적 이용, 그리고 병해충 발생예측 및 확산방지 등의 기후변화에 직접적으로 대응할 수 있는 대책을 마련하고 있지만, 국제 곡물가격 상승에 의한 간접적 위험성은 충분히 고려하지 않고 있음을 확인하였다.

〈표 6〉 「국가 기후변화 적응대책(2011~2015)」 분야별 대책

분야	적응대책	관련 사회위험 이슈	비고
건강	- 취약계층 중심의 폭염피해 방지대책 마련 - 전염병·대기오염 감시 및 예·경보체계 강화, 대응매뉴얼 보급, 대기오염 관리 강화 - 알레르기 모니터링 강화 및 대응시설 확대	13, 112	부분별 적응 대책
재난·재해	- 기후변화에 따른 취약지역 분석 및 방재기준 강화 추진 - 재해위험시설 보수, 방재정보 전달체계 구축, 도시하수도시설 개선 - 기후 친화적 국토이용·관리체계 구축 및 도시의 기후변화 적응능력 제고	12	
농업	- 기후변화 적응 작물재배기술 및 신품종 개발·보급 추진 - 농업용수의 효율적 이용 및 안정적 공급방안 마련 - 풍수해 예방, 병해충·가축질병 발생예측 및 확산방지체계 구축	15	
산림	- 한반도 생물다양성 유지를 위한 산림생물 종·자원 현지 내·외 보전 - 지역·수종별 취약성 평가 및 대책추진으로 임업생산성 유지·증진 - 산불·산사태 및 병해충으로 인한 산림피해 방지대책 추진	15, 16	
해양·수산업	- 해수면 상승으로 인한 연안변화 관리체계 및 적응방안 마련 - 어장변화 감시·예측 및 미래수산자원 확보 추진 - 수산생물 감염성 질병·산성화 피해저감 및 수산업 재해경감 대책 수립	12, 19	
물관리	- 물관리 취약성 완화를 위한 기반조성 및 시설개선 사업 추진 - 4대강 살리기 사업을 통한 홍수대응능력 향상, 대체수원 확보 등 안정적 물공급 - 하천·호소 수질관리 강화 및 하천생태계 보전·복원	12, 17, 18, 110, 111, 112	
생태계	- 생태계 및 지표종 모니터링 강화 및 취약성 평가 실시 - 생물종 및 유전자원 보전·복원 및 한반도 생태계 연결사업 추진 - 외래종 및 돌발 대발생으로 인한 피해방지 및 관리대책 수립	16	

<표 6> 「국가 기후변화 적응대책(2011~2015)」 분야별 대책(계속)

분야	적응대책	관련 사회위험 이슈	비고
기후변화 감시예측	- 기후변화 현상 감시·예측 기술 선진화 및 표준 시나리오 생산 - 독자적인 지구시스템·지역기후 모델 개발 및 상세 기후정보 생산 - IT 등 신기술을 활용한 맞춤형, 실시간 기후정보 전달체계 구축		간접적 적응 대책
적응산업·에너지	- 산업 분야별 적응대책 수립 가이드라인 개발·보급 - 기후변화에 따른 적응 신사업 발굴·지원 및 일자리 창출 - 에너지 분야 취약성 평가 및 공급 안정성 확보		
교육홍보·국제협력	- 기후변화 현실과 적응 필요성 및 적응 메커니즘에 대한 시민·학생 교육 강화 - 기후변화 적응교육 프로그램 개발 및 정보전달체계 구축 - 적응분야 별로 선진·개도국 및 국제기구와 협력 강화		

※ 자료: 환경부(2010).

‘전력수급 계획’토픽의 사회위험 이슈는 탈핵추세와 그로 인한 화력발전소의 증설이다. 후쿠시마 원전사고 이후 세계는 원전의 안전성을 이유로 탈핵추세로 가고 있는 실정이고, 그 예로 독일은 2011년 5월 말 원자력으로부터 ‘단계적 탈출’을 선언하고 “가능한 한 이른 시일 내에 재생에너지 시대에 도달하는 것이 목표”임을 분명히 한 바 있다. 우리나라의 경우 제6차 전력수급기본계획에서 원전에 대한 국민 수용성이 낮아진 상황을 이유로 원전 건설 확대계획 대신 석탄 화력발전소 12기를 건설토록 허용하였다. 석탄은 기후변화를 일으키는 온실가스 배출의 주범이고, 발전소에서 발생한 연기는 다이옥신, 질소산화물, 황산화물 등을 포함하고 있으며 이는 모든 생명체의 생존을 위협한다. 이러한 석탄발전의 환경문제와 원전의 안정성 문제는 현재까지 많은 논란 속에 중요한 국내 사회 위험요인으로 고려되어지고 있다.

V. 결론

사회의 급격한 변화와 사회를 구성하는 다양한 시스템 및 구성요소간의 격차와 한계 초과로 인하여 최근 많은 인적·사회적 재난이 발생하고 있다. 이에 전통적으로 기술예측 분야에 집중하여 적용되던 미래연구를 사회 위험 및 재난분야에 접목하는 사례가 증가하고 있다. 사회 위험을 분석하기 위해서는 사회 현상의 변화나 주요 쟁점 사안을 파악해야 하지만, 이는 정성적 데이터이기 때문에 전문가의 주관적인 의견을 보완할 수 있는 객관적 데이터가 요구되어 진다. 따라서 본 연구는 미래연구, 특히 미래 이슈를 발굴하는 환경스캐닝 분야에서 전문가의 판단을 뒷받침할 수 있는 근거를 제시하는 방법을 제안하는 것을 목표로 하였다.

우선 미래 위험이 될 수 있는 이슈를 효율적으로 파악하기 위하여 본 연구에서는 WEF에서 제시한 글로벌 리스크를 주제로 뉴스를 추출하여 국내 세부 리스크를 좁혀나가는 방식으로 접근하였다. 발생

가능성과 파급영향력이 큰 상위 50개 리스크 중 국내 영향력이 가장 높다고 판단되는 ‘기후변화 적응 실패’와 ‘수자원 공급위기’를 대상 리스크로 선정하였으며, 이와 관련된 뉴스 기사를 추출하였다. 추출된 뉴스 기사를 텍스트마이닝하여 토픽을 분석한 결과, ‘기후변화 적응실패’문서군과 ‘수자원 공급위기’ 문서군의 주요 토픽은 각각 6개로 분석되었고, 토픽을 구성하는 핵심 키워드와 연관 키워드, 가중치, 원문검토를 통해 각 토픽으로 유발될 수 있는 사회위험 이슈를 도출하였다. 도출된 사회위험 이슈를 국가 기후변화 적응대책과 비교한 결과, 부분별 적응대책에 사회위험 이슈가 모두 상응하는 반면, 간접적인 적응기반 대책은 매칭되지 못하는 것으로 나타났다. 이는 뉴스 기사를 분석 대상으로 하였다는 자료적 한계로 인한 것으로 사료되고, 이를 보완하기 위해서는 도출된 사회위험 이슈를 기반으로 전문가 회의 및 설문조사를 실시하는 등의 추가적 연구가 수행되어야 할 것이다.

또한 대부분의 국가 기후변화 적응대책이 본 연구에서 도출된 사회위험 이슈와 매칭되는 것을 확인하였으나, 국가 적응대책이 주로 직접적 대응 및 준비에 초점이 맞춰져 있고 타 분야와의 연계성과 이로 인한 촉발 위험성에 대해서는 인지와 대비가 부족한 것으로 판단된다. 그러므로 텍스트마이닝 등의 빅데이터 분석은 각 분야별 관점에서의 대응에서 벗어나 연계성과 협업이 가능해 질 수 있는 관점의 공유를 제공할 수 있을 것으로 기대되고, 이러한 과정을 통해 총체적 위험에 대비해 나갈 수 있을 것이다.

다만, 주제와 관련된 뉴스 기사를 선별하는 과정에서 주관성을 감소시켜야 할 필요성이 있다. 어떤 키워드 군집을 대상 주제로 볼 것인가, 즉 해당 주제의 정의와 범위에 따라 수집되는 뉴스 기사 방향성은 크게 달라질 수 있다. 본 연구에서는 관련문헌 검토와 브레인스토밍을 통하여 이 키워드를 선정하였으나, 추후 더욱 객관적인 방법론을 적용한다면 더 설득력 있는 결과를 도출할 수 있을 것으로 판단된다.

참고문헌

- 김성태. 2007. 또다른 미래를 향하여: 국정관리를 위한 미래예측과 미래전략, 경기: 법문사.
- 김원재. 2003. 한국사회 위험(Risk)의 특성과 치유. 사회연구. 5: 143-167.
- 오석원. 2011. 텍스트 마이닝의 실제적 활용에 관한 고찰. 고려대학교 석사학위논문.
- 울리히 백 지음, 홍성태 옮김. 2006. 위험사회: 새로운 근대(성)를 향하여. 서울: 새물결.
- 이영재, 이성수. 2011. 텍스트마이닝 기반의 인적재난사고사례 신뢰도 측정 연구. 정보시스템 연구. 20(3): 63-79.
- 이태봉, 이춘주. 2010. S&T Text Mining을 이용한 국방 유망기술 식별에 관한 연구. 한국국방경영분석학회지. 36(1): 39-49.
- 소방방재청. 2005. 사회안전지표 개발을 위한 국민안전의식조사.

- 정다미, 김재석, 김기남, 허종욱, 온병원, 강미정. 2013. 사회문제 해결형 기술수요 발굴을 위한 키워드 추출 시스템 제안. *지능정보연구*. 19(3): 1-20.
- 정영미. 2005. *정보검색연구*. 서울: 구미무역 출판부.
- 최진호, 김희수, 임남규. 2011. 기술예측을 위한 특허 키워드 네트워크 분석. *지능정보연구*. 17(4): 227-240.
- 한국과학기술기획평가원. 2011. 텍스트마이닝과 네트워크 분석을 활용한 미래예측 방법 연구.
- 한국보건사회연구원. 2013. 위험사회에 대한 국민의식조사.
- 한국보건산업진흥원. 2012. 국제곡물가격 상승에 따른 2013년 국내 식품물가 전망. *보건산업브리프*. 47: 1-8.
- 한국행정연구원. 2013. 소셜미디어를 통한 위험정보 확산과 위험인식에 관한 연구.
- 홍성만. 2013. 위험사회와 공공성 탐색: 불산가스 및 방사능 누출 위험사례를 중심으로. *한국정책연구*. 13(2): 117-262.
- 환경부. 2010. 국가 기후변화 적응대책(2011~2015).
- Beck, Ulrich. 1992. *Risk Society: Towards a New Modernity*. London: SAGE Publications.
- Evangelopoulos, Nicholas, Xiaoni Zhang, and Victor R. Prybutok. 2012. Latent Semantic Analysis: Five Methodological Recommendations. *European Journal of Information System*, 21(1): 70-86.
- Les Jones. 1993. Barriers to Creativity and Their Relationship to Individual, Group, and Organizational Behavior. *Nurturing and Developing Creativity: The Emergence of a Discipline*. 136.
- Salton, G., A. Wong, and C. S. Yang. 1975. A Vector Space Model for Information Retrieval. *Journal of the American Society for Information Science*. 18(11): 613-620.
- Spark, Jones, K. 1972. A Statistical Interpretation of Term Specificity and Its Application in Retrieval. *Journal of Documentation*. 28(1): 1-21.
- WORLD ECONOMIC FORUM. 2012. 2012년 글로벌 리스크 보고서(제7판).

원진영: 수자원공학 석사학위(논문: BREAH 모형을 이용한 fill dam의 붕괴유출특성 분석)를 취득하고, 현재 국립 재난안전연구원 선임연구원으로 재직 중이다. 현재 미래재난 예측과 관련된 연구를 수행하고 있으며, 관심분야는 사회위험, 미래예측, 텍스트마이닝, 토양 방재 등이다. 주요 논문으로는 “RUSLE 모형의 산지유역 적용을 위한 지표관리 인자 산정식 제안(2012)”, “강우에너지식 유도를 통한 전국 월별 강우침식인자 특성 분석(2013)” 등이 있다(wjy30228@korea.kr).

김대관: 수문학 석사학위(논문: 하천환경 모니터링을 위한 취수보의 수문특성 분석), 수공학 박사학위(논문: SWAT 및 HEC-RAS 모형을 결합한 내성천 유역의 토사수지분석)를 취득하고, 현재 국립재난안전연구원 연구관으로 재직 중이다. 2011년 “소하천 설계기준 재설정”, 2012년 “재해예방사업 투자현황 비교분석”, 2013년 “복합적

미래예측 방법론 분석을 통한 미래재난예측 기법 개발” 연구를 수행하였고, 현재 미래재난 예측과 관련된 연구를 계속하고 있으며 최근 관심분야는 사회위험, 미래예측, 텍스트마이닝, 자연 재난(홍수, 집중호우 등) 등이다. 주요 논문으로는 “미래재난예측을 위한 R-Scanner 기술개발(2014)”, “텍스트마이닝을 활용한 수자원연구트렌드 분석(2014)”, “A study of Disaster case of Newspaper Articles using Topic detection and tracking analysis(2014)” 등이 있다(kdgon1609@korea.kr).