

데이터 분석 기반 미래 신기술의 사회적 위험 예측과 위험성 평가

서용운[†]

부경대학교 안전공학과

(2017. 2. 9. 접수 / 2017. 3. 11. 수정 / 2017. 5. 31. 채택)

Data Analytics for Social Risk Forecasting and Assessment of New Technology

Yongyoon Suh[†]

Department of Safety Engineering, Pukyong National University

(Received February 9, 2017 / Revised March 11, 2017 / Accepted May 31, 2017)

Abstract : A new technology has provided the nation, industry, society, and people with innovative and useful functions. National economy and society has been improved through this technology innovation. Despite the benefit of technology innovation, however, since technology society was sufficiently mature, the unintended side effect and negative impact of new technology on society and human beings has been highlighted. Thus, it is important to investigate a risk of new technology for the future society. Recently, the risks of the new technology are being suggested through a large amount of social data such as news articles and report contents. These data can be used as effective sources for quantitatively and systematically forecasting social risks of new technology. In this respect, this paper aims to propose a data-driven process for forecasting and assessing social risks of future new technology using the text mining, 4M(Man, Machine, Media, and Management) framework, and analytic hierarchy process (AHP). First, social risk factors are forecasted based on social risk keywords extracted by the text mining of documents containing social risk information of new technology. Second, the social risk keywords are classified into the 4M causes to identify the degree of risk causes. Finally, the AHP is applied to assess impact of social risk factors and 4M causes based on social risk keywords. The proposed approach is helpful for technology engineers, safety managers, and policy makers to consider social risks of new technology and their impact.

Key Words : risk forecasting, risk assessment, data analytics, new technology, text mining, AHP

1. 서론

새로운 기술 개발은 효과적이고 혁신적인 기능을 개인과 사회에 제공하며, 경제성장과 사회발전에 큰 기여를 해왔다. 많은 기업들과 연구기관들은 신기술 개발에 노력을 기울이며, 지속적인 수익을 추구하거나 사회 인프라를 구축하여 인간 삶 향상에 도움을 주고 있다. 과거의 기술혁신을 보면, 인간이 할 수 없는 일을 기술이 대체하거나 자원의 효율적 활용을 통해 에너지 문제를 해결하는 등 다양한 순기능을 제공하였다.

그러나 역사적으로, 기술은 의도치 않은 역기능(unintended side effect)도 함께 존재해왔다¹⁻²⁾. 실제로 지금까지 기술진보의 순기능에만 초점을 맞추어 많은 부분 역기능을 무시하는 경향이 존재하였다. 예를 들어, 자동차의 개발은 교통사고라는 위험을 가져오고,

IT 기술의 발전은 개인정보를 침해했으며, 원자력발전소는 환경을 오염시킬 수 있는 잠재원인으로 지목되고 있다. 따라서 기술개발 수준이 성숙되고, 인간 삶의 진보가 어느 정도 만족하게 이루어진 현재 시점에서, 경제성장과 사회발전을 위해 역기능을 제어하는 방식이 되돌아 볼 필요가 있다³⁾. 성숙되고 안전한 사회구현을 위해서 언제까지 순기능의 긍정적인 효과에만 초점을 맞출 수는 없게 된 것이다. 2017년 1월에 열린 다보스포럼에서도 인공지능, 3D 프린터, VR 등 12가지 기술에 대한 미래 기술위험사회의 부정적인 결과를 경고하기도 했다⁴⁾.

이를 위해서, 앞으로 개발될 미래 신기술의 사회적 위험을 미리 예측하고 제거하기 위해 위험성을 평가하는 작업은 매우 중요하다⁵⁾. 그러나 아직까지 사회적 위험예측은 전문가의 평가에 의존하고 있으며, 안전공학

[†] Corresponding Author : Yongyoon Suh, Tel : +82-51-629-6467, E-mail : ysuh@pknu.ac.kr
Department of Safety Engineering, Pukyong National University, 45 Yongso-ro, Nam-gu, Busan 48513, Korea

과 관련한 프레임워크의 활용이 부족하다. 또한, 대부분 경제, 국방, 보안, 사회안전 측면에서 시나리오를 제공하고 있지만⁶⁻⁷⁾, 구체적으로 요인 수준에서 제시하는 부분은 부족하며, 특히 위험성을 정량적으로 평가하는 체계도 제공되지 않고 있다.

이와 같이 구체적인 요인 수준의 정량적 분석을 위해 본 연구에서는 미래 신기술의 위험성을 보고하거나 예측하는 문서 데이터를 활용하고자 한다. 최근 신기술의 사회적 위험성을 미리 예측하는 결과물로서 정책보고서 혹은 뉴스, 신문기사들이 다양하고 유용한 콘텐츠를 제공하고 있다. 국외의 경우 맥킨지(McKinsey), IDC, 가트너(Gartner) 등 컨설팅 자료나 외국 방송사인 CNN, IT 기술 리뷰 사이트인 CNET, 그 외 다수의 블로그나 전문 웹사이트에서 신기술에 대한 기능이나 사회적 위험성을 다루고 있다. 또한, 국내의 경우 한국과학기술기획평가원(KISTEP)이나 과학기술정책연구원(STEPI)에서 정책적으로 기술영향 보고서를 연구하는 경우가 많다. 또한, 전자신문이나 디지털타임즈와 같은 전문적인 신문사들도 과학기술과 관련한 사회이슈들을 많이 보도하고 있다. 최근, 국내의 경우에도 국립재난안전연구원에서 소셜미디어로 재난이슈를 분석하는 연구를 수행하고 있다⁸⁾.

이를 반영하여 본 연구의 수행단계는 크게 세 가지로 이루어진다. 첫째, 미래 신기술과 관련한 사회적 위험을 추출하기 위해 기술보고서, 뉴스, 컨설팅 보고서 등으로부터 텍스트마이닝을 활용하여 키워드를 추출한다. 둘째, 키워드로 도출된 사회적 위험요인을 안전관리의 위험원인분석에서 사용되는 4M(Man, Machine, Media, Management)의 위험원인⁹⁾에 분류한다. 마지막으로, 4M의 세부원인이 사회적 위험성에 미치는 영향력을 분류된 텍스트마이닝에서 도출된 위험요인에 근거하여 계층분석과정(AHP: Analytic Hierarchy Process)을 활용해 위험성을 평가한다¹⁰⁾. 결과적으로 다양한 위험요인 키워드를 담고 있는 문서 데이터를 활용하여 위험요인을 예측하고 위험원인을 구조화하여 평가하는 일련의 데이터 분석 기반 위험성 평가 체계를 구축할 것이다.

2. 신기술의 사회적 위험성

신기술은 아직 사회에 직접 적용되고 응용되지 않은 상태로 기술적이고 공학적인 위험성보다도 사회에 적용됐을 경우의 위험성을 예측하기 어렵다¹⁾. 특히, 기술적으로는 공학자 간의 합의하에 문제해결방법을 발견하기가 비교적 용이하지만, 사회적 위험성은 각계 분야의 사회과학자마다 제각기 제시하는 위험요인이 다

Table 1. Factors of STEEP frame

STEEP	Factors
Society	Health, Population, Life-style, Safety, Security, Operations, Culture
Technology	R&D, Automation, Innovation, Product, Diffusion, Breakthrough, Productivity
Economics	Economic growth, interest rate, inflation, GDP, Consumer
Ecology	Weather, Climate, Pollution, Ecosystem, Maintenance, Prevention, Physiology
Policy	Law, Regulation, Government, Institute, Regime, National environment

를 수 있다⁴⁾. 또한, 기술을 개발하는 사람들은 기술 자체의 오작동이나 구현불가능성을 위험으로 고려하지만, 사회과학자들은 사회와 환경, 인간과의 연관관계에서 그 위험을 찾는다는 점에서 큰 차이가 있다³⁾. 환경과학자들은 환경에 대한 사회적 위험성을 고려할 수 있고, 인간공학자들은 사용 시 사용자들에게 영향을 미칠 수 있는 위험성을 제시할 수 있다. 또한, 도시공학자들은 기존에 구축된 도시 인프라들과 기술의 영향관계에서 불안 요인들을 발견할 수 있다. 이처럼 사회적 위험성은 각계에서 미리 예측하지만 서로 합의가 어렵고, 토의하기 힘든 단점이 있다. 즉, 기술개발 시 과학자나 공학자들의 집단지성(collective intelligence)이 이루어지기 힘들 가능성이 높다. 따라서 신기술에 대한 다양한 분야의 사회과학자들의 사회적 위험성을 통합하여 위험성을 분석할 필요가 있다.

최근 미래 거시환경 분석 과정에서 STEEP 방법론이 활용되고 있는 것도 이와 같은 기술자, 사회과학자, 환경과학자들의 의견을 다양하게 반영하기 위해서이다¹¹⁾. STEEP는 사회(Society), 기술(Technology), 경제(Economics), 환경(Ecology), 정책(Policy)을 동시에 반영하는 프레임으로 Table 1과 같이 일반적인 기준을 제시할 수 있다. 본 연구에서는 “사회적”의 의미를 넓게 사용하는 차원에서 사회, 환경, 정책과 관련된 요인들을 주요 요인으로 포함한다. 이 요인들은 데이터 분석을 통해 도출된 위험 키워드들을 클러스터링 결과의 그룹들을 분석하고 정의할 때 활용한다.

3. 데이터 기반 위험성 평가 방법론

3.1 연구방법론 구조

본 연구는 미래 신기술의 사회적 위험성을 서술하거나 정리한 웹 문서들의 텍스트 데이터에 기반하여 Fig. 1과 같이 미래 신기술의 사회적 위험성 평가를 공학적이고 정량적인 방법으로 제시한다. 우선 첫 번째 단계

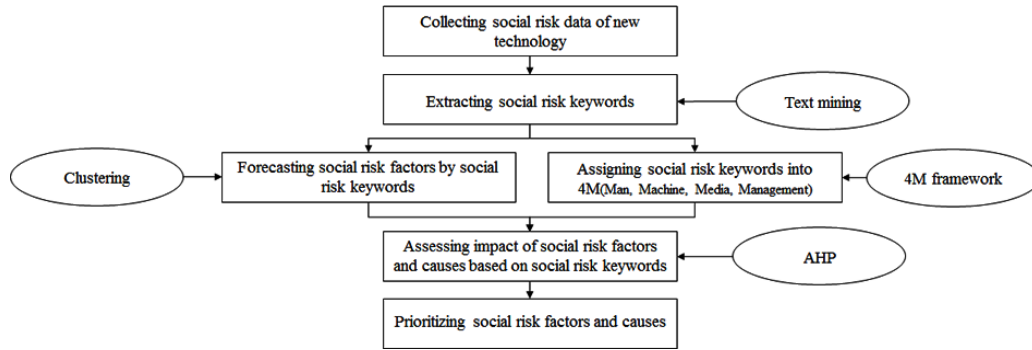


Fig. 1. Data-driven social risk assessment of new technology.

로 앞에서 제시한 사회적 위험성 콘텐츠를 포함하거나 포함할 가능성이 높은 정책보고서와 신문기사와 같은 비정형 데이터를 수집한다. 두 번째로 수집한 정책보고서와 신문뉴스의 키워드를 텍스트마이닝을 사용하여 도출하여, 도출된 키워드를 클러스터링을 통해 위험요인으로 예측하고, 위험원인을 4M 분류에 맞춰 제시 한다. 마지막으로 도출된 AHP를 사용하여 위험원인들에 대한 위험요인의 사회적 위험성을 평가한다.

3.2 데이터 수집

위험성 평가를 위해 신기술의 대상을 선정하여 관련 문서를 수집한다. 관련 문서인 정책보고서나 신문기사 등은 pdf 파일이나 워드 파일, 또는 웹 HTML 문서로 제공된다. 이들은 특정 보안문서를 제외하고는 모두 텍스트 파일로 변환 가능하며, 텍스트 파일은 모두 R 을 이용한 분석이 가능하다. 문서의 수집은 자동적으로 수집하는 crawling 코드를 사용할 수도 있고, 수동으로 직접 수집할 수도 있다. 본 연구는 일부 HTML 문서는 파이썬(Python) 기반의 crawling 코드를 이용하여 자동으로 수집하였고, 자동으로 수집하기 어려운 pdf 나 워드 형태의 문서는 수동으로 수집하였다.

3.3 위험요인 예측 : 텍스트마이닝

텍스트마이닝은 텍스트로 이루어진 문서형태의 비정형 데이터를 분석하는 대표적인 방법으로 초기 자연어 처리의 한 방법이라고 볼 수 있다¹²⁾. 문서에 포함된 키워드의 빈도수와 동시발생 가능성을 포함하여 다양하게 분석될 수 있다. 일반적으로 텍스트마이닝의 과정은 크게 세 가지로 구분된다. 첫 번째로 텍스트를 하나의 집합체인 Corpus로 인식하여, 이를 클래스로 추출하는 parsing 과정을 거친다. 두 번째로 텍스트 중에서 불용어(stop words) 처리를 하게 된다. 일반적으로 의미가 없는 관사, 전치사, 조사 등의 품사를 삭제하고, 영어의 경우 stemming이라 불리는 관련어 제거 처리를

하게 된다. 예를 들어, 동사의 현재형과 과거형, 미래형을 동일하게 현재형으로만 키워드를 추출하는 방식이다. 마지막으로 각 문서에서 나온 키워드의 수를 토대로 문서×키워드 행렬을 도출하게 된다. 신기술의 사회적 위험성을 포함한다고 고려되는 문서들을 텍스트마이닝을 통해 사회적 위험 키워드를 추출하면, 가능한 사회적 위험요인들을 예측할 수 있다.

3.4 위험원인 제시 : 4M 프레임워크

4M은 위험원인을 체계적으로 고려하기 위해 제공된 형식으로 인간(Man), 기계(Machine), 매체(Media), 관리(Management) 차원에 대하여 안전 관리자로 하여금 대책을 수립할 수 있도록 지원한다⁹⁾. 본 연구에서는 사회적 위험 키워드를 통해 4M 측면에서 위험원인을 제시하여, 4M 중 가장 위험성이 높은 원인을 개략적으로 제시하고 확인한다.

3.5 위험요인 및 원인의 위험성 평가 : AHP

다양한 위험요인 평가 기법 중 신기술에 대한 예측 위험요인은 데이터의 부족으로 인해 순수한 정량기법으로 평가하기 어렵다. 따라서 정성적이면서도 정량적인 분석이 가능한 AHP 분석을 활용하고자 한다. AHP는 목적을 평가하기 위한 기준요인들을 계층화시켜 쌍대비교(pairwise comparison)한다는 점에서 보다 체계적으로 분석목표를 평가할 수 있는 방법론으로 위험성 평가에 주로 사용되고 있다¹³⁻¹⁴⁾. 본 연구에서는 사회적 위험 키워드들을 바탕으로 사회적 위험원인과 위험요인을 AHP로 평가한다.

4. 사례 분석 : 드론의 사회적 위험성 평가

본 연구에서는 신기술과 관련하여 드론을 사례분석으로 제시한다. 드론은 현재 국방, 물류, 재난 등에 활용방안이 활발히 모색 중이며, 우리나라 역시 최근 국

가 미래성장동력 중 하나로 선정되기도 하였다. 미래부에서는 드론의 활용도를 높이기 위해, 드론 제품을 허가대상에서 신고대상으로 그 의무성을 낮추었다. 또한, 비면허 대역에서의 주파수 활용폭도 기존 2.4 GHz 외에 다양한 대역대에서 자유롭게 사용할 수 있도록 허가하였다. 한편으로는 드론에 대한 기능적, 사회적, 정책적 문제점을 고려하기 위해 법적차원에서 이를 규제하는 연구도 수행되고 있다¹⁵⁾. 그러나 아직까지 드론에 대한 사회적 역기능과 위험성을 공학적인 방법으로 다룬 연구는 찾아보기 힘들다. 따라서 본 연구에서 제시하는 데이터 기반 사회적 위험예측과 위험성 평가에 활용하기 적합한 미래 신기술로 고려하였다.

4.1 드론 관련 데이터 수집

드론 관련 데이터를 수집하기 위해 구글에서 “Drone”, “Social”, “Society”, “Risk”로 검색하면 약 90 만개(2016년 10월 10일 기준)의 게시글이 Fig. 2와 같이 검색되었다. 이 중에서 본 연구에서 제시하는 방법론의 적합성을 사례분석하는 차원에서 1페이지에서 20 페이지까지 200개의 게시물을 웹 crawling을 이용하거나 수동으로 직접 수집하였다.

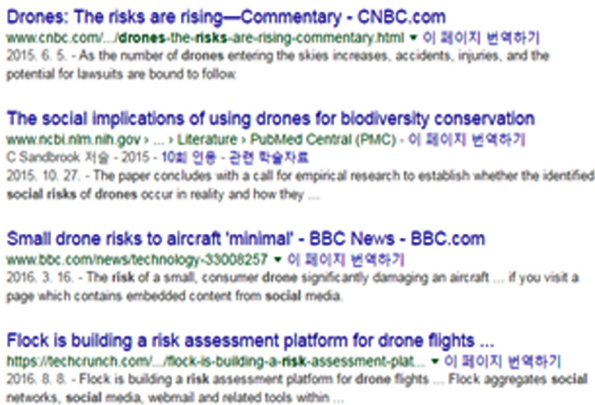


Fig. 2. Web-documents searched from Google web site.

4.2 드론의 사회적 위험 요인 예측

위에서 수집한 200개의 문서 중 일부 사용 불가능한 자료(웹 호환 불가, 텍스트 깨짐 현상)를 제외하고 약 132개의 문서를 R 프로그램을 사용하여 텍스트마이닝을 수행한 결과, 키워드는 500개 이상이 도출되었다. Table 2는 도출된 사회적 위험 키워드 중의 일부를 제시한 표이다.

그 중 문서에 나온 키워드가 나왔는지 여부의 binary 빈도 수(문서에 나왔으면 빈도에 관련없이 1, 나오지 않으면 0)를 이용하여, 빈도수 상위 50개의 키워드로 132

Table 2. Social risk keywords extracted from textmining

Social risk keyword
bother, privacy, dangerous, breaking, park, animal, nano, movement, data, armed, avoid, wifi, technology, biomimetic, sound, location, safety, scare, camera, news, apartment, arrest, annoy, detect, film, public, hotel, place, small, photo, violation, radar, moral, home, upset, habits, journalism, celebrities, accident, enforcement, live, intercept, surveillance, drugs, police, criminal, video, intercept, infrastructure, risk, national security, illegal, transmission, device, hacker, policy, activity, abuse, location, safety, accident, aircraft, crash, training, injure, property, regulation, kill, emergency, sky, visual, fall, condition, instruction, hit, hazard, life, environment, interfere, enforcement, detect, operation, noise, transmission, network, tracking, sound, hazard, alert, detection, mission, disrupt, task, poacher, hide, service, infrastructure, armed, attack, terror, weapon, threat, moral, bomb, security, surveillance

Table 3. Category of social risk factors extracted from keyword

Social risk factor	Social risk keyword
Operation	Collision Control, Flight, Building, Telegraph pole, Vehicle, Fall, Crash, Discharge, Pathway, Pedestrian
	Pollution Noise, Appearance, Aircraft, Air(Sky), Shadow, Delivery
Security	Hacking Hacker, Intercept, Illegal use, Credit card, Delivery, Personal information, Vulnerability*
	Privacy Camera, Photo, Personal information, Peeping, Small, Video, Dwelling, Crime
National Defence	Terrorism Attack, Military, Illegal remodeling, Weapon, Threat, Small, Crime, Public

* Vulnerability는 보안이 약한 것을 의미

개의 문서를 k-means 클러스터링 분석하였다. 클러스터링 개수 k의 기준은 CCC(Cubic Clustering Criterion)를 활용하여 CCC 값이 최고가 되는 5개로 선택하였으며, Table 3와 같이 구조화하였다. 결과적으로 5개의 위험요인을 정성적으로 다시 세 가지의 위험요인으로 정리하였다. 첫째, 운영적(Operation) 위험요인으로 충돌(Collision)과 공해(Pollution)를 포함한다. 둘째, 보안(Security)도 사회적 위험요인으로 해킹과 사생활과 관련한 위험요인들이 도출되었다. 셋째, 국방(National defence)과 관련하여 테러요인들 역시 사회적 위험으로 다양한 키워드들이 추출되었다.

4.3 드론의 사회적 위험 원인 제시

드론의 사회적 위험요인으로 추출된 분류들은 위험성의 원인분류 체계를 따르지 않는다는 한계가 있다. 위험원인은 크게 인적, 물적, 그리고 관리적 측면으로 구분되며, 4M의 경우 이에 대해 체계적으로 나눌 수 있는 틀을 제공하고 있다. 따라서 4M에 맞추어 사회적 위험 키워드들을 분류하여, 위험원인을 제시한다.

Table 4는 앞에서 도출된 사회적 위험 키워드들을

Table 4. Categorization of social risks into 4M causes

4M	Risk cause	Social risk keyword
Man	Unsafe behavior	Control, Fall, Discharge, Delivery
	(Wicked) Misuse	Attack, Peeping, Weapon, Threat, Crime, Hacker
Machine	Malfunction	Fall, Discharge, Control, Crash
	Poor design	Small, Illegal remodeling, Vulnerability
Media	Place	Building, Dwelling, Public
	Impact material	Building, Telegraph pole, Vehicle, Pedestrian, Aircraft
	Information	Camera, Photo, Personal information, Movie, Credit card
Management	Lack of Regulation	Flight, Noise, Appearance, Air, Shadow, Public, Hacker, Intercept

통해 4M에 포함되는 위험원인들을 제시하였다. 먼저, 인간(Man)과 관련한 재해원인으로는 불안정한(unsafe) 행동과 오남용(misuse)을 들 수 있으며, 이에 따라 드론의 제어나 방전, 낙하나 드론을 사용한 공격, 도촬, 위협 등의 위험요인들이 발생하게 된다. 다음으로, 기계(Machine)와 관련해서 드론의 기능불량(malfunction)에 따라 발생하는 낙하, 방전, 파괴나 잘못된 설계(poor design)로 인한 불법개조, 보안취약성이나 발견하기 어려운 정도의 초소형 드론이 위험원인으로 제시될 수 있다. 또한, 매체(Media)의 경우는 드론의 사용장소(place), 충격매체(impact material), 정보매체(information)가 위험원인으로 지목되고 있다. 기본적으로 건물이나 주거지, 공공에서 무단으로 활용될 수 있으며, 그 장소에서 건물, 전봇대(telegraph pole), 자동차, 보행자, 비행기 등 접촉하여 사고를 일으킬 수 있다. 이 외에도 정보매체 차원에서 드론에 카메라를 달아 도촬하여 사진이나 동영상 정보로 저장한다거나, 개인정보나 신용카드 정보 등을 드론 네트워크를 통해 훔쳐올 수 있다. 마지막으로 관리차원에서는 대부분 규제부족(lack of regulations)과 관련된 부분으로 무단비행이나 소음, 외견, 그림자 등의 하드웨어 차원의 규제나 해킹, 도난과 같이 소프트웨어 차원의 규제도 요구되고 있다.

4.4 드론의 사회적 위험성 평가

마지막으로, 드론의 위험요인을 위험원인 측면에서 AHP를 통해 평가하여, 사회적 위험성을 제시한다. 위에서 분류된 위험원인들이 사회적 위험성에 미치는 영향과, 각 위험원인들이 사회적 위험요인에 미치는 영향을 고려하여 각각의 원인들과 사회적 위험요인의 가중치를 도출한다. 특히, 데이터 기반 사회적 위험성 평가는 사회적 위험 키워드들을 위험원인에 제시하므로

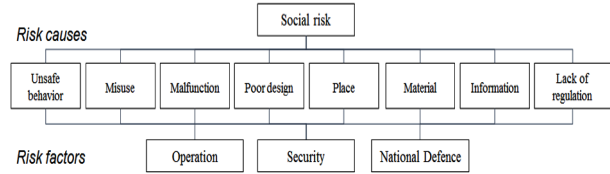


Fig. 3. AHP model for social risk assessment.

써, 실질적인 가이드라인 없이 위험원인을 평가했던 기존 연구의 한계점을 극복할 수 있다는 강점이 있다. 즉, 불안정한 행동과 오남용을 단순히 위험원인의 용어로만 평가하는 것이 아니라, 불안정한 행동의 사회적 위험 키워드와 오남용의 사회적 위험 키워드를 데이터 분석 결과로 제시하여 보다 객관적인 평가를 수행할 수 있다. 예를 들어, 드론 추락, 방전, 배달 사고들로 인한 불안정한 행동이 물리적 공격, 사생활 촬영, 도난 등으로 인한 오남용보다 어느 정도 위험할지에 대한 구체적 근거를 제시하고 평가할 수 있는 것이다. 이에 따른 Fig. 3는 위험원인들이 위험요인에 미치는 영향을 평가하기 위한 AHP 모델을 나타낸다.

Table 5는 AHP의 평가결과를 나타낸다. 평가는 드론 기술에 관련한 ICT 연구자, ICT 기술경제·정책 평가 전문가, 안전공학 전문가 등 다양한 분야의 전문가 20명에게 의뢰하였다. 특히, ICT 관련 대부분은 한국전자통신연구원(ETRI) 소속의 전문가들에게 의뢰하여 드론의 최신 기술 및 정책방향이 반영되도록 하였다. AHP의 집단평가(group decision) 방법을 위해 기하평균법을 사용했으며, 소프트웨어로는 Expert Choice 11 version을 활용하였다.

그 결과, 기본적으로 위험원인들 중에 가장 큰 영향력을 미치는 원인은 정보매체(0.318)와 오남용(0.208)으로 나타났다. 드론이나 관리체제 보다는 인간의 오용

Table 5. Weight of each element and alternative

Risk causes		Weight of risk alternatives		
		Operation	Security	National Defence
Man	Unsafe behavior (0.024)	0.634(1)	0.174(3)	0.192(2)
	Misuse (0.208)	0.105(3)	0.637(1)	0.258(2)
Machine	Malfunction (0.048)	0.481(1)	0.114(3)	0.405(2)
	Poor design (0.043)	0.143(3)	0.429(1)	0.429(1)
Media	Place (0.063)	0.172(2)	0.726(1)	0.102(3)
	Impact material (0.117)	0.364(2)	0.099(3)	0.537(1)
	Information (0.318)	0.084(3)	0.705(1)	0.211(2)
Management	Lack of Regulation (0.180)	0.433(2)	0.466(1)	0.100(3)
Total		0.246(3)	0.499(1)	0.255(2)

이나 드론이 처리하고 수집할 수 있는 정보매체들이 미래 사회에 위험원인들로 지목되었다. 각 위험요인들의 위험원인들과의 관련성을 살펴보면, 운영적 위험성의 경우는 불안정한 행동(0.634), 기능불량(0.481)에 가장 큰 영향을 받는 것으로 나타났다. 보안 위험성은 오남용(0.637), 잘못된 설계(0.429), 장소(0.726), 정보매체(0.705), 규제부족(0.466)에 기인한 위험요인으로 드러났다. 마지막으로 국방 위험성은 잘못된 설계(0.429), 충격매체(0.537)의 위험원인으로부터 도출된다고 분석되었다. 결과적으로 전체적으로는 보안 위험성이 미래 드론 사용에 있어서 가장 심각한 위험요인이며, 이에 대한 미래 사회대책이 필요하다.

5. 결론

본 연구는 미래 신기술이 도입될 때 사회에 미칠 위험성에 대한 보고서 및 신문기사 등의 텍스트 데이터 분석을 통해 보다 알고리즘과 정량화에 기반한 위험예측 및 위험성 평가를 수행하였다. 기존에 일부 한정된 설문 위주의 분석이 아닌 다양한 전문가와 의견 제시자의 전자·웹 문서 데이터를 수집하여 집단지성 기반의 위험요인을 예측하는데 기여하였다. 또한 위험요인 예측을 위해 사용한 비정형 문서 데이터 분석이 가능한 텍스트마이닝 방법에 더해 위험원인 분석을 위해 전통적인 안전관리기법 중 하나인 4M을 활용하였다. 이를 통해 텍스트마이닝의 결과인 위험 키워드들을 사고원인 4M에 맞추어 어떤 위험원인들이 제시될 수 있는지 확인하였다. 기존의 4M의 프레임에 주고 전문가가 직접 위험원인을 적은 것이 아닌 데이터에서 추출된 유의미한 위험요인 키워드들에서 위험원인을 도출하였다. 이 결과는 안전관리자에게 해석을 용이하게 하고, 통찰력을 줄 수 있을 것으로 기대된다. 또한 AHP 평가를 수행하여 위험요인들과 원인들의 영향력을 평가하였다. 본 분석 방법론의 최대 기여점 중 하나는 AHP를 통해 위험원인과 위험요인을 평가할 때, 데이터 분석으로 도출된 사회적 위험 키워드들을 가이드라인으로 제공하여 중요도를 근거에 기반해 객관적으로 평가할 수 있도록 지원했다는 점이다. 이를 통해 도출된 사회적 위험요인과 위험원인의 중요도에 따라 본 연구에서는 오남용과 정보매체의 위험원인을 관리하고, 정보보안과 관련한 위험요인을 제거해야한다는 결과를 제시하였다.

이와 같은 기여점에도 불구하고 본 연구는 데이터 기반의 위험성 평가를 위한 초기 연구로서 향후 연구가 필요하다. 첫째, 데이터의 부족이다. 텍스트 데이터

분석은 기본적으로 대용량의 빅데이터를 기반으로 해야 보다 양질의 키워드를 추출할 수 있다. 그러나 본 연구는 제안연구의 사례분석 차원으로 200개 정도의 텍스트 문서만을 이용해 위험요인을 예측하였다. 따라서 후속연구로 정책보고서나 신문기사 외의 보다 많은 사회요인 텍스트를 포함한 데이터 유형을 활용하고, 고차원의 정보 처리 알고리즘을 활용하여 다양한 사회적 위험 요인 키워드를 도출해야 한다. 두 번째로는 4M 프레임워크의 확장이다. 본 연구는 가장 널리 그리고 유용하게 쓰이는 위험원인 분석 방법론 중 하나인 4M에 의존하여 수행되었다. 그러나 최근에는 4M 이외의 더 다양한 근본적 원인들을 제시하는 틀이 제시되고 있다. 특히 4M은 생산현장의 위험원인분석을 수행하는 전통적인 프레임으로 4M 이외의 원인분류를 보다 확장할 필요가 있다. 이처럼 위험원인을 4M 보다 세분화하면 사회적 위험 키워드를 통해 구체적인 위험원인을 제시하기에 용이할 것이다. 마지막으로, AHP 평가 방법에 있어, 키워드의 빈도를 활용하는 데이터 기반의 보다 정량적인 평가 방법을 개발하는 것도 향후 연구로 고려된다. 텍스트마이닝으로 도출된 사회적 위험 키워드 빈도수나 다른 키워드 지표(동시단어 빈도 등)를 활용하여 위험요인을 평가한다면, 전문가의 AHP 수행과는 다른 결과와 함께 부가적인 평가비교 정보를 제공하리라 기대된다.

감사의 글: 이 논문은 부경대학교 자율창의학술연구비(2016년)에 의하여 연구되었음

References

- 1) A. R. Kohler and C. Som, "Risk Preventative Innovation Strategies for Emerging Technologies: The Cases of Nano-textiles and Smart-textiles", *Technovation*, Vol. 34, No. 8, pp. 420-430, 2014.
- 2) R. E. Alcock and J. Bursby, "Risk Migration and Scientific Advance: The Case of Flame-Retardant Compounds", *Risk Analysis*, Vol. 26, No. 2, pp. 369-381, 2006.
- 3) T. Assmuth, M. Hilden and C. Benighaus, "Integrated Risk Assessment and Risk Governance as Socio-political Phenomena: A Synthetic View of the Challenges", *Science of the Total Environment*, Vol. 408, pp. 3943-3953, 2010.
- 4) World Economic Forum, "The Global Risks Report 2017-12th edition", World Economic Forum, 2017.
- 5) European Environment Agency, "Late Lessons from Early Warnings: The precautionary Principle", Copenhagen, Denmark, pp. 1896-2000, 2001.

- 6) M. Amer, T. U. Daim and A. Jetter, “A Review of Scenario Planning”, *Futures*, Vol. 46, pp. 23-40, 2013.
- 7) D. Sarpong and M. Maclean, “Scenario Thinking: A Practice-Based Approach for the Identification of Opportunities for Innovation”, *Futures*, Vol. 43, No. 10, pp. 1154-1163, 2011.
- 8) S. H. Choi, “The Detection Model of Disaster Issues based on the Risk Degree of Social Media Contents”, *Journal of the Korean Society of Safety*, Vol. 31, No. 6, pp. 121-128, 2016.
- 9) KOSHA, “4M Risk Assessment Technique”, 2011.
- 10) T. L. Saaty, “The Analytic Hierarchy Process”, McGraw-Hill, New York, NY, 1980.
- 11) C. Durst, M. Durst, T. Kolonko, A. Neef and F. Greif, “A Holistic Approach to Strategic Foresight: A Foresight Support System for the German Federal Armed Forces”, *Technological Forecasting and Social Change*, Vol. 97, No. 3, pp. 91-104, 2015.
- 12) R. Kostoff, D. Toothman, H. Eberhart and J. Humenik, “Textmining Using Database Tomography and Bibliometrics: A Review”, *Technological Forecasting and Social Change*, Vol. 68, No. 3, pp. 223-252, 2001.
- 13) A. Kokangul, U. Polat and C. Dagsuyu, “A New Approximation for Risk Assessment Using the AHP and Fine Kinney Methodologies”, *Safety Science*, Vol. 91, pp.24-32, 2017.
- 14) J. Lee, I. Han and S. R. Chang, “A Study on the Risk Level of Work Types in Urban Railway Construction”, *Journal of the Korean Society of Safety*, Vol. 31, No. 2, pp. 98-103, 2016.
- 15) N. J. Kim, S. E. Lee and H. J. Kim, “Related Laws and Performance Criteria for Public Service Drones for Disaster Safety”, *Journal of the Korean Society of Safety*, Vol. 31, No. 4, pp. 150-155, 2016.