

한국의 사회적 이슈 도출을 위한 뉴스 빅데이터 분석 연구

Analysis of News Big Data for Deriving Social Issues in Korea

이홍주(Hong Joo Lee)*

초 록

복잡해지고 있는 현대 사회의 뉴스 키워드를 시간적 흐름에 따른 빈도수와 상관관계로 분석하는 것은 이슈들에 대한 대응과 해결 방안을 논의하기 위해 매우 중요한 연구라고 할 수 있다. 이에 본 논문에서는 10년(2009~2018)간의 뉴스 빅데이터 분석을 통해 사회적 키워드의 흐름과 주요 이슈들 간의 관계를 분석하였다. 분석결과 본 연구에서는 정치적 이슈, 교육·사회문화, 젠더갈등 그리고 사회적 사건이 주요 이슈들로 제시되었다. 또한, 본 연구에서는 이슈의 변화와 흐름을 연구하기 위해 이를 5년 기준으로 양분하여 변화하는 것을 분석하였다. 이를 통해 사회적 이슈의 시간에 따른 변화와 그 대응방안을 연구하였다. 그 결과 국민생활과 밀접한 키워드(경제, 경찰)는 시간의 흐름에 관계없이 우리 사회에서 매우 중요하게 논의되는 키워드로 분석되었다. 또한 ‘안전’과 같은 키워드는 최근 들어 빈도수에 비해 증가율이 감소되었다. 이를 통해, 우리 사회가 안전에 대한 인식을 개선할 필요가 있는 것으로 추론할 수 있다.

ABSTRACT

Analyzing the frequency and correlation of the news keywords in the modern society that are becoming complicated according to the time flow is a very important research to discuss the response and solution to issues. This paper analyzed the relationship between the flow of social keyword and major issues through the analysis of news big data for 10 years (2009~2018). In this study, political issues, education and social culture, gender conflicts and social problems were presented as major issues. And, to study the change and flow of issues, it analyzed the change of the issue by dividing it into five years. Through this, the changes and countermeasures of social issues were studied. As a result, the keywords (economy, police) that are closely related to the people's life were analyzed as keywords that are very important in our society regardless of the flow of time. In addition, keyword such as 'safety' have decreased in increasing rate compared to frequency in recent years. Through this, it can be inferred that it is necessary to improve the awareness of safety in our society.

키워드 : 소셜이슈, 빅데이터, 미래 신호 탐색, 텍스트 마이닝, 키워드 이슈 프레임워크
Social Issues, Big Data, Signal Detection, Text Mining, Keyword Issue Framework

본 논문은 2016학년도 경기대학교 연구년 수혜로 연구되었음.

* Department of Industrial and Management Engineering, Kyonggi Univrsity(blue1024@kgu.ac.kr)

Received: 2019-08-05, Review completed: 2019-08-19, Accepted: 2019-08-22

1. 서 론

오늘날 현대사회의 복잡성과, 불확실성이 높아지고 있다. 과학기술의 발달로 전 세계의 수많은 요소들이 상호간 작용하며, 하나로 통합된 시스템으로 움직이고 있다.

따라서, 현재의 우리사회의 이슈들을 연구하고 논의하는 일은 중요하지만, 이를 분석하는 것은 어려운 일이 되어 가고 있다. 왜냐하면, 우리사회의 사회적 이슈들의 변동성과 흐름을 모두 고려하기는 어렵기 때문이다. 그러므로, 현재의 흐름을 바탕으로 논의해야 할 사회적 이슈에 대한 연구는 중요하다고 할 수 있다. 그러나 정부 및 기업, 사회적 계층간의 위치에 따라, 그리고 각 집단과 개인의 가치관과 미래 비전에 따라 다양한 사회적 이슈들에 대한 판단은 달라 질 수 있다. 그럼에도 불구하고, 사회적 흐름에 따른 이슈 분석은 반드시 연구가 요구되는 분야라고 할 수 있다. 이에, 정부 및 각 기관에서는 다양한 사회문제 및 이슈를 논의하고 전문가들과 함께 그 흐름을 연구하고 있다. 현재, 한국사회는 성장 동력의 실종, 저출산 고령화의 도래, 경제사회 양극화, 일자리 부족 등의 다양한 사회문제를 겪고 있다[14].

또한, 주요 선진국들은 미래변화의 트렌드를 파악하고 미래의 핵심기술을 선별하기 위하여, 주기적으로 국가의 미래 트렌드를 분석하고 그 결과를 발표하고 있다[13]. 이에 각 연구기관들은 SNS를 비롯한 온라인 채널에서 생산되는 텍스트 형태의 비정형 데이터가 우리사회에서 발생하는 문제들을 설명하고 있는 경우가 많은 사실을 바탕으로, 소셜 네트워크 서비스 (social network service, SNS) 및 다양한 인터넷 사이트의 웹 크롤링(web crawling)을 통해

빅 데이터(Big Data)를 확보하고 이를 분석하는 연구를 수행하고 있다. 빅 데이터(Big Data)란 기존 데이터베이스 관리도구로 데이터를 수집, 저장, 관리, 분석할 수 있는 역량을 넘어서는 대량의 정형 또는 비정형 데이터 집합 및 이러한 데이터로부터 가치를 추출하고 결과를 분석하는 기술을 의미 한다[18]. 그러나, 수집기술과 분석기술의 어려움으로 연구의 한계가 존재하는 실정이다. 미국 전략비즈니스인사이트(SBI, 2011), 영국 Horizon 스캐닝센터(BIS, 2011), 핀란드 퓨처스연구센터(TrendWiki, 2011)등 많은 연구 그룹이 빅 데이터 분석을 통한 데이터 흐름의 신호를 모니터링 하기 위한 프로세스와 도구를 개발하였다[20, 3, 21]. 그러나, 이들의 접근방식은 주로 전문가들의 지식과 의견을 기반으로 미래를 전망하는 방법을 사용하여 왔다 [22]. 따라서, 이러한 방법은 관련분야의 다양한 전문가의 의견을 수렴해야 하는 어려움이 있다. 즉, 연구의 한계를 벗어나 다양한 데이터 분석을 기반으로 한 핵심이슈 선정과 이슈에 대한 트렌드 분석의 접근방법이 요구된다고 할 수 있다. 또한, 발생한 사회적 이슈들간에 어떤 관계가 있는가에 대한 연구를 통해 관계를 입체적으로 분석할 필요가 있다. 이에 본 연구에서는 이러한 문제를 연구하기 위해 아래와 같은 연구 질문을 제기하였다.

- ① 어떻게 사회적 이슈를 정형화된 방법으로 분석할 것인가?
- ② 어떻게 사회적 이슈를 시간의 흐름에 따라 분석할 것인가?
- ③ 사회적 이슈들은 어떠한 관계적 특성을 갖고 있는가?

이러한 연구 질문을 해결하기 위해, 본 연구에서는 10년(2009년~2018년) 동안 한국에서 발행된 뉴스매체의 인터넷 뉴스(사회면)를 통해 데이터를 수집하였다. 그리고, 뉴스의 비정형 데이터의 빈도와 증가율 분석을 통해 이슈의 흐름과 전망을 분석하였다. 아울러, 10년 동안의 사회적 이슈의 발생빈도를 바탕으로 요인분석을 통해 잠재요인들간의 관계를 분석하였다.

2. 이론적 배경

2.1 미래신호 연구

미래의 환경변화를 감지하기 위한 다양한 연구가 시도되고 있으며, 여러 연구 중에서 가장 많은 주목을 받고 있는 것은 미래의 변화를 예감할 수 있는 약신호(weak signal)를 탐지하는 것이다[23, 16]. 약신호는 ‘미래에 가능한 변화의 징후’[2]로 시간이 흐르면서 강신호(strong signal)로, 강신호는 다시 트렌드(trend)나 메가트렌드(mega trend)로 발전할 수 있다. Hiltunen (2008)은 약신호를 미래신호(future sign)라는 개념을 이용하여 *interpretation*, *signal*, *issue*와 같이 3차원의 미래신호 공간으로 설명하였다 [11].

2.2 약신호(Weak Signal) 연구

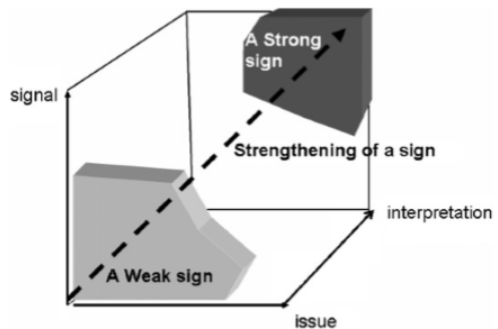
미래연구에서는 약신호를 미래의 추세, 변화, 신흥 현상 뒤에 있는 미래지향적인 정보라고 정의하고 있다[23]. 따라서 약신호는 현재에 영향을 미치지 않지만, 향후 추세와 시스템

의 불연속성을 이끄는 지표로 작용할 수 있다.

Mojjanen(2003)는 약신호는 변화를 유발하고, 미래변화의 신호임을 연구하였다 [15].

Ansoff(1975)는 전략적 계획의 일부분제를 해결하기 위해 약신호를 연구하였다[2]. 또한, 그는 약신호를 아직 불완전하여, 정확하게 추정하기 힘든 사건이라고 설명하였다[1].

Coffman(1997) 또한, 약신호에 대해 오랫동안 연구하였다[4-8]. 그는 정보이론, 사이버네트워크, 복잡성, 조직을 분석하는 방법으로 약신호를 연구하였다. 이를 통해 사업 환경에서 약신호를 사용하는 실용적인 측면에 중점을 두어 연구하였다. 그밖에 다양한 연구자들 역시 비즈니스의 환경과 기업조직을 약신호를 통해 연구 하였다. Hiltunen(2008)은 ‘미래신호’를 미래의 다양한 환경변화를 예측하는 데 핵심적인 것으로 연구하였다[11]. 그는 <Figure 1>과 같이 *signal*, *issue*, 그리고 *interpretation*의 세 가지 요인 통해 미래신호를 분석할 수 있는 개념을 제시하였다.



Source: Hiltunen(2008).

<Figure 1> Strengthening of the Future Sign

<Figure 1>과 같은 3차원 공간에서는 약신호에서 강신호(Strong Signal)로의 변화를 통

해 미래의 징후를 파악할 수 있다. 따라서, 약신호에 대한 연구는 기업의 전략적 의사결정에 도움을 줄 수 있다. 이에 많은 연구자들이 약신호에 대해 연구하였다. Piriene(2010)은 기업의 비즈니스 전략을 수립하는데 도움을 주기 위해 약신호를 연구해야 한다고 제시하였다[17].

또한, Hiltunen(2008)은 기업의 비즈니스 환경분석을 통해 직원들의 미래 지향적 사고를 유도하고, 비즈니스 환경 및 조직을 분석하기 위해 약한신호가 활용될 수 있음을 연구하였다[11]. 아울러 다양한 연구자들 역시 비즈니스 환경분석과 조직을 분석하기 위해 약신호를 연구하였다[9, 12, 18]. 이러한 접근법을 바탕으로 약신호 분석은 불확실한 미래 사업 분석의 한 방법으로 많은 주목을 받아 왔다.

3. 데이터 수집 및 분석방법

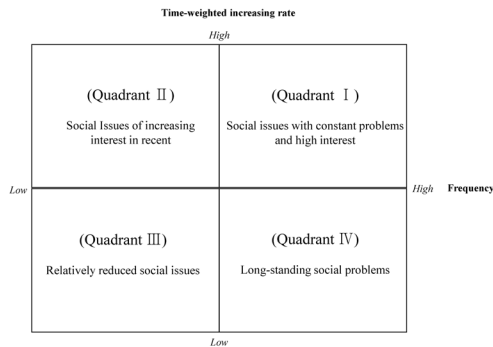
Downs(1972)의 연구에 따르면, 사회적 이슈는 사회적 관심을 받고, 절정기에 이르고 쇠퇴와 소멸하는 다섯 단계가 존재한다고 제시하였다[10]. 이에 많은 후속 연구들은 언론 보도량을 측정해 이슈들의 발전-생존 주기를 밝히고자 했다[11, 12]. 이에, 본 연구에서는 미래신호를 분석하기 위해 과거 이슈가 되었던 신문기사의 키워드를 시계열 흐름으로 분석하였다. 주어진 데이터에서 키워드의 발생 빈도는 일부 한계를 내포하고 있으나, 일반적으로 키워드의 중요도를 측정하는 방법으로 간주된다[19]. 이에 본 연구에서는 키워드의 발생빈도와 시간이 가중치 된 증가율을 통해 키워드 발생과 흐름의 특성을 연구하였다.

3.1 데이터 수집방법

서론에서 언급했던바와 같이, 본 연구에서는 2009년부터 2018년까지 한국에서 발행된 신문의 뉴스(사회면)를 수집하였다. 특히, 특정 키워드를 입력하고 뉴스를 검색하는 방법이 아닌, 한국에서 발표된 모든 언론사의 인터넷 뉴스(사회면)의 모든 기사를 수집하였다. 이를 위해 본 논문에서는 텍스트 마이닝(Text mining) 방법을 데이터 분석에 활용하였다. 텍스트 마이닝은 비정형 형태의 텍스트 데이터 에서 의미 있는 새로운 정보를 추출하는 것을 목적으로 하는 빅 데이터 분석방법이다. 이것은 문서에서 표현된 단어들을 바탕으로 내용을 분석하여 작성자의 의도를 파악할 수 있다. 또한, 문서 분류 및 군집화, 개념 요소 추출, 요소 관계 모델, 감정 분석, 문서 요약 등에 주로 활용되고 있다. 본 논문에서는 텍스트 데이터를 수집하기 위한 방법으로 R 프로그램을 활용하여, 웹 크롤링(Web crawling)을 통해 데이터를 확보하였다. 수집된 데이터를 바탕으로 광고, 숫자, 구어체, 오타 및 문장부호 와 같이 뉴스와 관련 없는 단어들을 처리하기 위한 ‘불용어 사전’을 구축하였다. 아울러, 형태소 분석을 통해, 데이터 중 주요 한글 품사를 바탕으로 단어를 추출하였다. 이를 바탕으로 데이터 분석 특성에 맞는 ‘전용사전’을 별도로 구축해, 분석에서 의미를 갖는 단어가 누락되지 않도록 전처리 과정을 진행하였다. 또한 데이터 수집 후 분석을 위해 빈도수를 기준으로 상위 100개의 키워드를 선별하였다. 추출된 키워드의 총 빈도수는 10,823,531개 이었다. 그러나 10,823,531번 도출된 단어를 분석하고 표현하기에는 한계가 있어, 본 연구에서는 발생빈도수에 따라 이를 다시 35개의 핵심키워드로 분류하였다.

3.2 분석방법

본 연구에서는 Hiltunen(2008)의 연구를 기초로 Yoon(2012)이 연구한 KIM(Keyword Issue Map)을 참고하였다[11, 23]. 또한 BCG의 제품포트폴리오 매트릭스를 연구한 Henderson(1982)의 연구를 참고하여 키워드 분석을 위한 프레임워크를 <Figure 2>와 같이 제시하였다[10].



<Figure 2> Keyword Issue Framework

본 연구에서는 <Figure 2>의 프레임워크를 완성하기 위해 ‘키워드 빈도수’와 ‘증가율’이라는 2개의 기준을 제시하였다. 이를 프레임워크의 X축(평균단어빈도)과 Y축(평균증가율)으로 나타냈다. 아울러, 중위수(median)로 분면을 나누는 기준으로 사용하였다. <Figure 2>에서 1사분면은 빈도수와 증가율이 모두 증가하고 있는 사회적 키워드가 위치하며, 이는 중요한 사회적 이슈 신호로 분석할 수 있다. 2사분면은 발생되고 있는 빈도수에 비해 증가율이 상대적으로 높은 영역으로 최근 활발하게 등장하는 사회적 이슈들이 분포하는 영역으로 분석할 수 있다. 3사분면은 발생빈도와 증가율이 상대적으로 낮은 키워드가 분포한다. 만약, 시간의 흐름에 따라 1, 2, 4 사분면에서 3사 분면으로 키워

드가 이동했다면 키워드의 중요도가 상대적으로 줄어든 것으로 볼 수 있다. 그러나, 새롭게 등장한 키워드라면 아직 크게 두드러지지 않은 잠재된 신호로 분석할 수 있다. 또한, 4사분면에 위치한 키워드는 단어빈도 수는 높으나 증가율이 높지 않은 키워드들로 우리사회에서 만성화된 이슈들과, 중요하지만 자칫 놓칠 수 있는 중요한 이슈들이 키워드들이 분포한다. 이에 본 연구에선 식 (1), 식 (2)를 바탕으로 <Figure 2>를 완성하였다. 식 (1)은 수집된 키워드의 평균 빈도를 계산한 방법이다. 아울러 본 연구에서는 단어들의 평균빈도를 계산할 때 시간에 따른 가중치 0.05를 부여하였다[23].

$$Average\ frequency = \frac{\sum_{t=1}^j [CF_t \times \{1 - tw(n-j)\}]}{tn} \quad (1)$$

$$Increasing\ rate = \frac{CF_t - CF_{t-1}}{CF_{t-1}} \times \{1 - tw(n-j)\} \quad (2)$$

CF_t = t시점에 추출된 단어의 빈도 수,
 CF_{t-1} = t시점 직전에 추출된 단어의 빈도 수,
 tn = 데이터가 수집된 기간 횟수, tw = 0.05,
 n = 전체시간구간, j = 현재년도

본 논문에서는 2009년부터 2018년 까지 한국에서 발행된 신문의 인터넷 뉴스 사회면의 자료를 수집하였고, 핵심 키워드를 도출하였다. 아울러, 매분기마다 수집된 키워드들의 시간에 따른 가중치(0.05)를 부여하여 키워드의 빈도를 계산하였다. 또한 도출된 키워드의 분기별 증가율을 각각 계산하였다. 그러나 특정년도의 일부기간에서 분야별 키워드 수집자료의 오류가 발생하여 본 연구결과에서는 이를 누락하였

다. 아울러, 분석을 위해 본 연구에서는 SPSS 24.0을 활용하였다.

4. 연구결과

4.1 키워드 분석

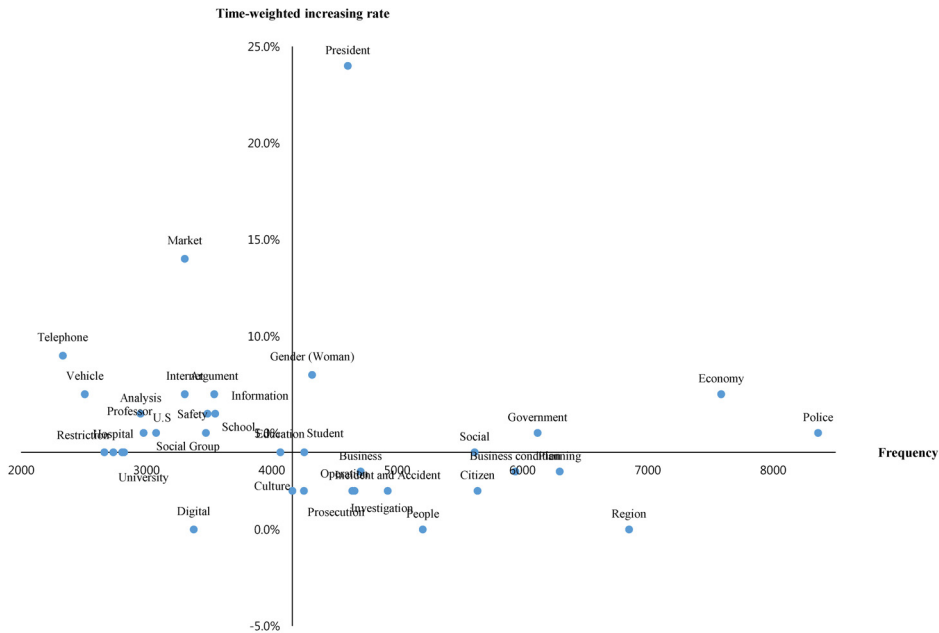
본 논문에서는 시간에 따른 가중치가 반영된 핵심 키워드를 1차 100개에서 2차 35개로 정리하고 그 결과를 <Table 1>부터 <Table 4>까지 나타냈다.

4.2 Keyword Issue Map

본 연구에서는 <Table 1>에서 <Table 4>까지의 분석자료를 바탕으로 Keyword Issue

Framework(KIF)을 활용하여 Keyword의 상대적 위치를 <Figure 3>, <Figure 4>와 같이 제시하였다. 아울러, 수집된 데이터를 5년 단위로 구분하여 시간의 흐름에 따른 키워드의 분석을 하였다.

<Figure 3>에서의 1사분면은 키워드의 빈도수와 증가율이 높은 영역으로 사회적으로 중요한 이슈가 되고 있는 키워드가 도출되는 영역이다. 본 연구에서 1사분면의 핵심키워드는 경찰, 경제, 정부, 여성, 대통령 이었다. 그러나 <Figure 4>에서의 1사분면 핵심키워드는 대통령, 경찰, 경제, 시민 사건 운영, 수사, 국민, 학생 등 이었다. 이는 해당기간 동안 발생한 한국의 정치적 사건(대통령 선거, 대통령 탄핵 등)을 고려했을 때, 나타날 수 있는 특수한 결과가 반영된 것으로 추정할 수 있다. 또한, 여전히 ‘경제’ 및 ‘경찰’은 상대적으로 매우 높은 빈도를 나타내고 있다.



<Figure 3> Keyword Position Map(2009~2013)

〈Table 2〉 Thirty-Five Keywords Collected Using Web Crawling (II)

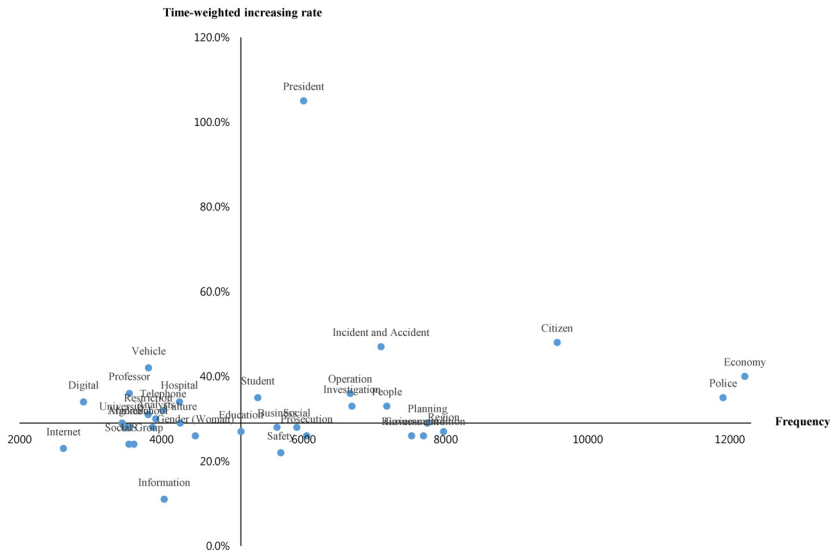
Weighted reflection	Social		Investigation		Operation		President		Business		Prosecution		Student		Education		Safety	
	Increasing Rate	Weighted reflection	Increasing Rate	Weighted reflection	Increasing Rate	Weighted reflection	Increasing Rate	Weighted reflection	Increasing Rate	Weighted reflection	Increasing Rate	Weighted reflection	Increasing Rate	Weighted reflection	Increasing Rate	Weighted reflection	Increasing Rate	Weighted reflection
5894		4854		3974		2798		4374		4728		2882		3364		2835		
6332	7%	5781	19%	4445	12%	16498	400%	4410	1%	6198	31%	3850	34%	4045	20%	3126	10%	
5767	-9%	3554	-39%	4590	3%	7746	-53%	4383	-1%	3078	-50%	3799	-1%	4157	3%	3170	1%	
5590	-3%	3539	0%	4830	5%	3421	-56%	5414	24%	3508	14%	4353	15%	4515	9%	3653	15%	
4905	-12%	3886	10%	4930	2%	2757	-19%	4684	-13%	3833	9%	3735	-14%	4667	3%	3186	-13%	
5126	5%	3275	-16%	4461	-10%	4052	47%	4828	3%	4179	9%	4090	10%	4490	-4%	3475	9%	
5285	3%	4524	38%	4630	4%	3331	-18%	5365	11%	3765	-10%	4463	9%	4684	4%	3869	11%	
4819	-9%	4586	1%	4506	-3%	3177	-5%	4610	-14%	4246	13%	4040	-9%	3570	-24%	3343	-14%	
4780	-1%	4442	-3%	4462	-1%	3246	2%	4154	-10%	3792	-11%	4133	2%	3578	0%	3828	14%	
6198	30%	5651	27%	4844	9%	3921	21%	5027	21%	4867	28%	5632	36%	4416	23%	3629	-5%	
4914	-21%	3884	-31%	4371	-10%	2654	-32%	4252	-15%	3797	-22%	4374	-22%	3917	-11%	3789	4%	
4694	-4%	4410	14%	3933	-10%	3132	18%	3882	-9%	3785	0%	3982	-9%	3263	-17%	2423	-36%	
4739	1%	5405	23%	4686	19%	4863	53%	4177	8%	4535	21%	5118	28%	3798	16%	3128	29%	
5079	7%	5709	6%	5083	8%	4030	-17%	4937	18%	4803	5%	4951	-3%	3886	2%	3202	2%	
5087	0%	4522	-21%	5042	-1%	3978	-1%	4544	-8%	3867	-19%	3724	-25%	4802	24%	3165	-1%	
3983	-22%	4054	-10%	3807	-24%	4372	10%	3538	-22%	3846	-1%	3249	-13%	2997	-38%	2397	-24%	
6594	66%	5689	40%	4933	30%	5119	17%	4687	32%	4837	26%	5387	66%	3988	33%	3562	49%	
7292	11%	6009	6%	5351	8%	4242	-17%	5709	22%	5056	5%	5212	-3%	4090	2%	5167	45%	
9670	33%	4760	-21%	5307	-1%	4187	-1%	6479	13%	4071	-19%	3920	-25%	5055	24%	5287	2%	
7366	-24%	5970	25%	6589	24%	3375	-19%	5310	-18%	5199	28%	4620	18%	5005	-1%	4488	-15%	
6853	-7%	8177	37%	7003	6%	4818	43%	4341	-18%	8352	61%	8882	92%	4452	-11%	8934	99%	
2217	-68%	4140	-49%	2695	-62%	1385	-71%	2215	-49%	3475	-58%	2406	-73%	2118	-52%	2882	-68%	
1597	-28%	1790	-57%	1767	-34%	1030	-26%	1674	-24%	1909	-45%	1310	-46%	1469	-28%	2715	-6%	
7943	397%	9087	408%	9483	437%	473	-54%	8329	397%	6769	253%	5674	333%	6646	343%	7506	176%	
6105	-23%	6492	-29%	6831	-28%	3106	557%	5844	-30%	5617	-17%	5596	-1%	5676	-15%	5985	-20%	
10003	64%	8908	33%	10984	61%	7420	139%	10269	76%	7791	39%	9078	62%	9525	68%	7494	25%	
3853	-61%	5964	-38%	5239	-52%	1733	-77%	4590	-55%	4852	-38%	3811	-58%	4173	-56%	4165	-44%	
5654	47%	8804	64%	6800	30%	16002	824%	6065	32%	9257	91%	6290	65%	4563	9%	5487	32%	
9358	66%	9418	7%	11692	72%	21860	37%	9820	62%	9472	2%	9692	54%	9155	101%	9163	67%	
2542	-73%	2725	-71%	2857	-76%	3156	-86%	2775	-72%	2764	-71%	2118	-78%	2365	-74%	2349	-74%	
4776	88%	5383	98%	5564	95%	6112	94%	4709	70%	5729	107%	4373	107%	4247	80%	5535	138%	
4235	-11%	5545	3%	5273	-5%	6447	5%	4251	-10%	4534	-21%	3607	-18%	3826	-10%	3968	-29%	

〈Table 3〉 Thirty-Five Keywords Collected Using Web Crawling (III)

Gender (Woman)	Culture		Information		School		Argument		Hospital		Market		Analysis		Social Group	
	Weighted reflection	Increasing Rate	Weighted reflection	Increasing Rate	Weighted reflection	Increasing Rate	Weighted reflection	Increasing Rate	Weighted reflection	Increasing Rate	Weighted reflection	Increasing Rate	Weighted reflection	Increasing Rate	Weighted reflection	Increasing Rate
2584	3609		2654		2597		2182		1990		1770		2141		2760	
3175	4842	34%	3144	18%	3002	16%	2004	-8%	2662	34%	2359	33%	2187	2%	3554	29%
3352	4900	6%	3319	6%	3394	13%	1834	-9%	3538	33%	1985	-16%	2323	6%	3025	-15%
3275	5363	9%	3438	4%	3794	12%	2360	29%	3185	-10%	2207	21%	2564	10%	3222	7%
3175	4308	-20%	3522	2%	3334	-12%	2293	-3%	2562	-20%	2230	-7%	2536	-1%	2892	-10%
3197	4806	12%	3665	4%	3273	-2%	2152	-6%	2443	-5%	2178	-2%	2818	11%	3356	16%
3843	4932	3%	3440	-6%	3620	11%	2824	31%	2534	4%	2767	27%	2710	-4%	2745	-18%
3738	4471	-9%	2833	-18%	2917	-19%	2914	3%	2772	9%	5147	86%	2387	-12%	3059	11%
3875	3434	-23%	3183	12%	3370	16%	3123	7%	2501	-10%	4804	-7%	3398	42%	2688	-12%
4671	4799	40%	3598	13%	3899	16%	3669	17%	2456	-2%	2823	-41%	3958	17%	3314	23%
4442	4287	-11%	3356	-7%	3230	-17%	2750	-25%	2826	15%	3381	20%	3291	-17%	3287	-1%
3821	3575	-17%	4203	25%	2825	-13%	2679	-3%	2563	-9%	6942	105%	2235	-32%	2748	-16%
3889	3302	-8%	5398	28%	4390	55%	4977	86%	2673	4%	5525	-20%	2276	2%	2721	-1%
6659	4808	46%	3496	-35%	3640	-17%	4703	-5%	2890	8%	2860	-48%	3066	35%	2977	9%
6026	3026	-37%	3011	-14%	3645	0%	4665	-1%	3098	7%	2354	-18%	2911	-5%	2733	-8%
4943	2936	-3%	2201	-27%	2784	-24%	4325	-7%	2153	-31%	1966	-16%	2601	-11%	1963	-28%
4094	3476	18%	3394	54%	4621	66%	5884	36%	2814	31%	5816	196%	3707	43%	3508	79%
7009	5061	46%	4219	24%	3832	-17%	7993	36%	3042	8%	3010	-48%	4259	15%	3415	-3%
6343	3185	-37%	5330	26%	3837	0%	3927	-51%	3261	7%	2478	-18%	4725	11%	4476	31%
4598	4306	35%	4828	-9%	4165	9%	3351	-15%	4674	43%	2823	14%	4167	-12%	3285	-27%
4546	3838	-11%	5440	13%	4128	-1%	5106	52%	5301	13%	2490	-12%	4469	7%	3736	14%
1894	1858	-52%	2098	-61%	1775	-57%	1467	-71%	1948	-63%	1041	-58%	2085	-53%	1418	-62%
1424	1471	-21%	2271	8%	1086	-39%	1426	-3%	1430	-27%	1231	18%	1285	-38%	1302	-8%
7174	5624	282%	7650	237%	4837	345%	5638	265%	7220	405%	5107	315%	6386	397%	4445	241%
5307	4519	-20%	3634	-53%	4350	-10%	3221	-43%	5514	-24%	3811	-25%	4202	-34%	3477	-22%
8538	8048	78%	4323	19%	6831	57%	6467	101%	7030	28%	6190	62%	6203	48%	6201	78%
3699	4581	-43%	4131	-4%	2973	-56%	2471	-62%	3101	-50%	2598	-58%	2918	-53%	2427	-61%
3631	3144	-31%	4589	11%	3751	26%	3560	44%	3866	25%	4201	62%	3840	32%	4053	67%
6638	7610	142%	7257	58%	6548	75%	4399	24%	5988	55%	5844	39%	5636	47%	6188	53%
2043	1778	-77%	1884	-74%	1638	-75%	1271	-71%	1515	-75%	1682	-71%	1416	-75%	1622	-74%
3490	3856	117%	2169	15%	3449	111%	2364	86%	3751	148%	2557	52%	3351	137%	3680	127%
3062	2936	-24%	1879	-13%	2671	-23%	2854	21%	2403	-36%	3081	20%	2816	-16%	3072	-17%

〈Table 4〉 Thirty-Five Keywords Collected Using Web Crawling (Ⅳ)

U.S		Digital		Restriction		Professor		University		Telephone		Vehicle		Internet	
Weighted reflection	Increasing Rate	Weighted reflection	Increasing Rate	Weighted reflection	Increasing Rate	Weighted reflection	Increasing Rate	Weighted reflection	Increasing Rate	Weighted reflection	Increasing Rate	Weighted reflection	Increasing Rate	Weighted reflection	Increasing Rate
2446		6965		2636		2096		2136		1981		1866		2635	
3608	48%	3109	-55%	2694	2%	2810	34%	2576	21%	1844	-7%	2192	17%	3088	17%
2839	-19%	3062	-1%	2055	-24%	2544	-9%	2607	1%	1870	1%	2300	5%	3109	1%
3098	5%	2741	-11%	2221	8%	2650	4%	3206	23%	2065	10%	2147	-7%	3010	-3%
2814	-9%	2791	2%	1828	-18%	2365	-11%	2827	-12%	2122	3%	2485	16%	2859	-5%
3377	20%	3066	10%	2055	12%	2757	17%	2615	-7%	1954	-8%	1841	-26%	2977	4%
3023	-10%	2907	-5%	2196	7%	2488	-9%	2840	9%	1804	-8%	2853	55%	2889	-3%
2770	-8%	2678	-8%	2100	-4%	2457	-2%	2535	-11%	1772	-2%	2172	-24%	3015	4%
2561	-8%	2774	4%	1901	-10%	2828	15%	2844	12%	1854	5%	2889	24%	2556	-15%
3235	26%	2942	6%	2728	44%	3605	27%	3900	37%	2069	12%	2219	-18%	3207	25%
2737	-15%	2859	-3%	2248	-18%	3109	-14%	2811	-28%	1715	-17%	2997	35%	2973	-7%
2803	2%	2622	-8%	2389	6%	2620	-16%	2786	-1%	2112	23%	1874	-37%	2821	-5%
2321	-17%	3229	23%	2615	9%	2516	-4%	2452	-12%	2070	-2%	2290	22%	3385	20%
3599	55%	4371	35%	3393	30%	3005	19%	2834	16%	3083	49%	2414	5%	6334	87%
3230	-10%	3651	-16%	3414	1%	3050	1%	3117	10%	2414	-22%	2596	8%	3525	-44%
2355	-27%	2539	-30%	2389	-30%	2494	-18%	2365	-24%	1817	-25%	1984	-24%	2199	-38%
2443	4%	3399	34%	3250	36%	2940	18%	2581	9%	2638	45%	2996	51%	3460	57%
3788	55%	4601	35%	5056	56%	3384	22%	2983	16%	3374	28%	3326	31%	3865	12%
3400	-10%	3843	-16%	3423	-32%	3661	2%	3281	10%	5733	70%	3796	-3%	4892	27%
3366	-1%	3706	-4%	3591	5%	3458	-6%	3663	12%	6173	8%	3474	-8%	3849	-21%
3018	-10%	2778	-25%	4325	20%	3711	7%	2606	-29%	6732	9%	3856	11%	4523	18%
1514	-50%	1730	-38%	1822	-58%	1495	-60%	1069	-59%	2157	-68%	1476	-62%	1645	-64%
1700	12%	1335	-23%	1254	-31%	871	-42%	1025	-4%	1395	-35%	922	-38%	1229	-25%
6202	265%	7151	436%	5878	369%	4572	425%	4530	342%	8571	514%	6346	588%	4485	265%
4448	-28%	3440	-52%	3764	-36%	3963	-13%	3789	-16%	5565	-35%	3485	-45%	2936	-35%
6565	48%	4894	42%	6069	61%	6851	73%	7506	98%	7118	28%	5750	65%	6153	110%
2354	-64%	1554	-68%	3228	-47%	2462	-64%	2741	-63%	2714	-62%	3734	-35%	1212	-80%
3082	31%	1921	24%	4132	28%	4238	73%	4458	63%	3549	31%	3678	-1%	2083	72%
5292	72%	4027	110%	6377	54%	6706	57%	6015	35%	3724	5%	6163	68%	3280	58%
1509	-71%	825	-80%	1437	-77%	1320	-80%	1391	-77%	943	-75%	1964	-68%	712	-78%
3798	152%	1796	118%	3536	146%	3000	127%	2679	93%	1954	107%	4023	105%	1596	124%
2429	-36%	1717	-4%	2650	-25%	2266	-24%	2200	-18%	1779	-9%	2746	-32%	867	-46%



〈Figure 4〉 Keyword Position Map (2014~2018)

이것은 사회적 사건과 경제적 이슈가 우리 사회에서 관심 있게 지켜보아야 하는 매우 강력한 사회적 신호로 분석할 수 있다. 특히 ‘경제’의 경우 뉴스에서 언급되는 빈도수가(2009~2013)년도에 비해 증가되었음을 알 수 있다. 이것은 ‘경제’ 문제에 대한 관심이 지속적으로 나타나고 있음을 의미한다. 또한 〈Figure 3〉, 〈Figure 4〉의 2사분면은 빈도수에 비해 상대적으로 증가율이 높게 상승하고 있는 키워드들이 나타나는 분면이다. 즉, 2사분면에 등장하는 키워드들은 최근 사회적 관심이 빠르게 증가하고 있는 키워드들이라 할 수 있다. 〈Figure 3〉에서는 시장, 디지털논란, 정보, 학교, 안전, 단체, 분석, 차량, 미국, 구속 등의 단어가 주요 키워드로 분석되었다. 반면 〈Figure 4〉에서는 차량, 교수, 전화, 병원, 구속, 인터넷 등이 주요 키워드로 분석되었다. 특히, ‘구속’의 경우 사회적 사건과 이를 처리하는 결과에 대한 관심이 지속적으로 연결되고 있음을 설명한다. 또한 ‘차량(vehicle)’과 같은

운송수단(현재의 운송수단 및 미래형 운송수단)에 대한 관심도가 계속 증가하고 있음이 분석되었다. 반면 ‘안전’의 빈도수는 증가하였으나, 증가율은 감소하였다. 또한, ‘인터넷’, ‘논란(인터넷 상에서의 혐오적 표현 등)’ 역시 〈Figure 3〉에서는 2사 분면이었으나 〈Figure 4〉에서는 4사분면으로 이동하였다. 이러한 흐름을 통해 우리 사회가 ‘안전’, ‘인터넷’, ‘논란’ 과 같은 키워드에 대해 만성화된 반응을 갖게 될 수 있음을 추론할 수 있다. 아울러, ‘국민’, ‘시민’ 과 같은 키워드가 시간이 지남에 따라 4사분면(Figure 3)에서 1사분면(Figure 4)으로 이동하였다. 이것은 최근 정치적·사회적 이슈가 활발하게 논의되면서, 우리사회의 중요한 키워드로 등장하고 있음을 의미한다. 또한 〈Figure 3〉에서 4사분면에 위치한 ‘사건’, ‘수사’가 〈Figure 4〉의 1사분면으로 이동하였다. 이것은 ‘사건’ 및 ‘범죄’와 같은 사회문제가 우리사회에서 또 다시 핵심 이슈로 등장하며 논의되고 있음을 추론할 수 있다.

4.3 뉴스 키워드의 발생 증가율 분석

본 논문에서는 두 기간(2009~2013, 2014~2018)에 따라 뉴스 키워드의 상대적 증가율을 <Figure 5>와 같이 분석하였다. 즉, X축을 2009-2013년 동안의 키워드 증가율, Y축을 2014-2018년 동안의 키워드 증가율로 정의하고, 키워드들의 상대적 위치를 분석하였다. 분석결과 차량(vehicle), 경제, 경찰, 전화 등과 같은 키워드가 데이터 수집기간(2009~2018) 동안 계속 증가하는 것으로 분석되었다. 특히, '차량', '전화기'는 지난 10년 동안 지속적으로 발생률이 증가하는 것으로 분석되었다. 따라서, 정치 및 사회적 이슈를 제외하면 앞으로도 우리 사회에서 지속적으로 관심을 야기 시키는 이슈가 될 수 있을 것으로 추론할 수 있다.

4.4 요인분석

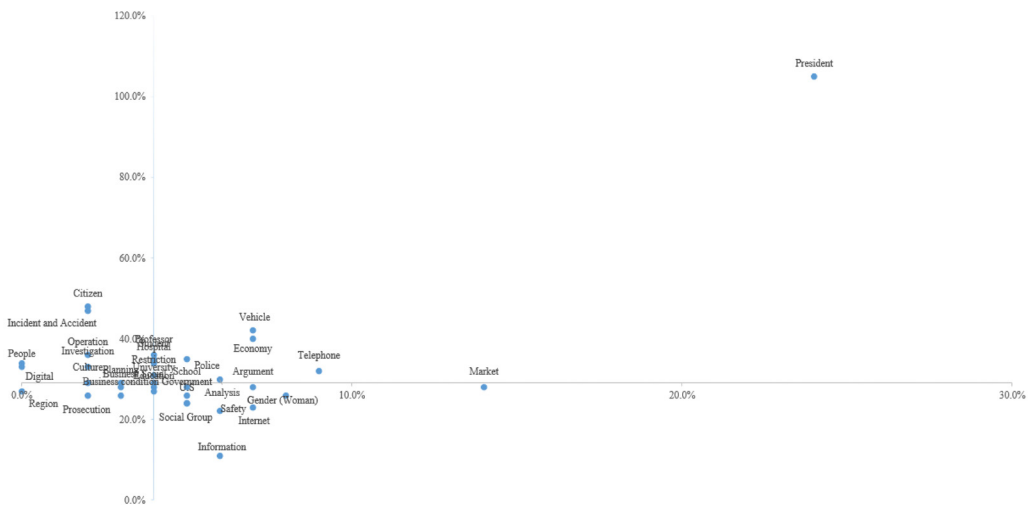
본 연구에서는 빅데이터 분석을 통해 도출된

키워드들의 잠재된 속성을 연구하고, 속성들의 연관성을 분석하였다. 이를 위해 본 연구에서는 최대우도법이 아닌 주성분 분석(Principal Component Analysis)과 베리맥스(Varimax with Kaiser Normalization) 방법을 활용하여 요인분석을 하였다. 본 연구에서는 <Table 5>를 통해 본 요인 분석이 타당함을 검증하였다.

<Table 5> KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy		0.889
Bartlett's Test of Sphericity	Approx. Chi-Square	5406.133
	df	528
	Sig.	0.000

또한, <Table 6>의 결과를 통해 군집화된 요인들의 Cronbach's Alpha 값은 모두 0.6 이상이였다. 따라서, 군집된 요인들의 신뢰성이 있는 것으로 분석되었다.



<Figure 5> Analysis of Keyword Growth Rate

아울러, 본 논문에서는 <Table 6>의 분석을 바탕으로, 35개의 키워드를 4개의 잠재요인(사회적 사건, 교육문화, 양성평등, 정치적 이슈)로 제시하였다.

또한 <Table 6>의 연구결과를 바탕으로, 4개의 잠재요인들간의 상관관계를 분석하였으며 그 결과는 <Table 7>과 같다. 이를 통해 4개의

잠재요인들간의 상관관계가 존재함을 분석하였으며, 또한 잠재요인들간의 방향성을 분석하였다. 이에 본 연구에서는 ‘사회적 사건’을 종속변수로, ‘교육문화’, ‘젠더갈등’ 그리고 ‘정치적 이슈’를 각각 독립변수로 하여 다중회귀분석을 수행하였다. 그 결과는 <Table 8>, <Table 9>와 같다.

<Table 9>를 통해, ‘교육문화’와 ‘젠더갈등’

<Table 6> Rotated Component Matrix

	1	2	3	4	Cronbach's Alpha
Telephone	0.843				0.975
Safety	0.824				
Hospital	0.790				
Incident	0.765				
Vehicle	0.758				
Investigation	0.746				
Police	0.733				
Analysis	0.725				
Restriction	0.720				
Economy	0.693				
People	0.677				
Operation	0.675				
Prosecution	0.674				
Business condition	0.638				
Planning	0.629				
Information	0.603				
Culture		0.814			0.968
Region		0.773			
Education		0.710			
Business		0.685			
United States of America		0.684			
University		0.634			
Argument			0.862		0.9
Internet			0.738		
Woman			0.717		0.818
President				0.875	
Citizen				0.781	

이 ‘사회적 사건’과의 인과적 관계가 있는 것으로 설명할 수 있다. 또한 ‘교육문화’가 ‘젠더갈등’ 보다 ‘사회적 사건’에 더 높은 계수를 나타내고 있는 것으로 제시되었다.

이에, 본 연구에서는 시간의 변화에 따른 뉴

스 키워드의 발생을 분석하기 위해, 기간(2009~2013, 2014~2018)에 따라 두 그룹으로 뉴스의 키워드 데이터를 각각 분류하였다. <Table 10>은 2009년부터 2013년도까지의 잠재요인들간의 상관관계 분석 결과이다.

<Table 7> Correlation Analysis(2009~2018)

		Social Problems	Education and Culture	Gender conflict	Political issue
Social Problems	Pearson Correlation	1	0.888**	0.717**	0.554**
	Sig.(2-tailed)		0.000	0.000	0.001
Education and Culture	Pearson Correlation	0.888**	1	0.660**	0.602**
	Sig.(2-tailed)	0.000		0.000	0.000
Gender conflict	Pearson Correlation	0.717**	0.660**	1	0.210
	Sig.(2-tailed)	0.000	0.000		0.250
Political issue	Pearson Correlation	0.554**	0.602**	0.210	1
	Sig.(2-tailed)	0.001	0.000	0.250	

* Correlation is significant at the 0.05 level (2-tailed).

** Correlation is significant at the 0.01 level (2-tailed).

<Table 8> Model Summary(2009~2018)

Model	R	R ²	Sig. F Change	Durbin-Watson
1	0.888 ^a	0.789	0.000	
2	0.905 ^b	0.807	0.035	1.817

^a. Predictors: (Constant), Education and Culture(2009~2018)

^b. Predictors: (Constant), Education and Culture(2009~2018), Gender equality problems(2009~2018)

^c. Dependent Variable: Social Problems(2009~2018).

<Table 9> Coefficients^a(2009~2018)

Model		Unstandardized B	Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	VIF
1	Constant	65.276	525.777		0.124	0.902	
	Education and Culture (2009~2018)	1.179	0.111	0.888	10.597	0.000	1.000
2	Constant	-226.539	511.889		-0.443	0.661	
	Education and Culture (2009~2018)	0.976	0.139	0.735	7.008	0.000	1.770
	Gender conflict(2009~2018)	0.334	0.151	0.232	2.213	0.035	1.770

^a. Dependent Variable: Social Problems(2009~2018).

<Table 11>, <Table 12>는 2009~2013년도의 잠재요인들간의 다중회귀분석 결과로, ‘젠더 갈등’과 ‘교육문화’가 ‘사회적 사건’에 영향을 미

치는 것으로 분석되었다.

<Table 13>은 2014년부터 2018년도까지의 잠재요인들간의 상관관계 분석 결과이다.

<Table 10> Correlation Analysis(2009~2013)

		Political issue	Education and Culture	Gender conflict	Social Problems
Political issue	Pearson Correlation	1	0.173	-0.134	0.220
	Sig.(2-tailed)		0.478	0.584	0.365
Education and Culture	Pearson Correlation	0.173	1	-0.004	0.376
	Sig.(2-tailed)	0.000		0.000	0.000
Gender conflict	Pearson Correlation	-0.134	-0.004	1	0.721**
	Sig.(2-tailed)	0.584	0.986		0.000
Social Problems	Pearson Correlation	0.220	0.376	0.721**	1
	Sig.(2-tailed)	0.365	0.112	0.000	

* Correlation is significant at the 0.05 level(2-tailed).

** Correlation is significant at the 0.01 level(2-tailed).

<Table 11> Model Summary(2009~2013)

Model	R	R ²	Sig. F Change	Durbin-Watson
1	0.721 ^a	0.520	0.000	
2	0.815 ^b	0.664	0.019	1.179

^a. Predictors: (Constant), Gender equality problems(2009~2013).

^b. Predictors: (Constant), Gender equality problems(2009~2013), Education and Culture(2009~2013).

^c. Dependent Variable: Social Problems(2009~2013).

<Table 12> Coefficients^a(2009~2013)

Model		Unstandardized B	Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	VIF
1	Constant	3155.424	366.688		8.605	0.000	
	Gender conflict(2009~2013)	0.406	0.095	0.721	4.291	0.000	1.000
2	Constant	1105.093	844.586		1.308	0.209	
	Gender conflict(2009~2013)	0.407	0.082	0.723	4.986	0.000	1.000
	Education and Culture (2009~2013)	0.480	0.183	0.379	2.618	0.019	1.000

^a. Dependent Variable: Social Problems(2009~2013).

<Table 14>, <Table 15> 역시 위와 같은 방법으로 2014에서 2018년 까지 수집된 데이터를 회귀분석한 결과이다. 그러나, <Table 10>, <Table 11>의 결과와는 달리 2013년에서 2018년

까지의 기간에는 ‘젠더갈등’과 ‘정치적 이슈’가 ‘사회적 사건’에 영향을 미치는 것으로 분석되었다. 또한, ‘젠더갈등’의 경우 데이터를 수집한 모든 기간에 걸쳐 ‘사회적 사건’에 영향을 주는 요인으

<Table 13> Correlation Analysis(2014~2018)

		Political issue	Gender conflict	Education and Culture	Social Problems
Political issue	Pearson Correlation	1	0.341	0.615*	0.538
	Sig.(2-tailed)		0.254	0.025	0.058
Gender conflict	Pearson Correlation	0.341	1	0.904**	0.950**
	Sig.(2-tailed)	0.254		0.000	0.000
Education and Culture	Pearson Correlation	0.615*	0.904**	1	0.948**
	Sig.(2-tailed)	0.025	0.000		0.000
Social Problems	Pearson Correlation	0.538	0.950**	0.948**	1
	Sig.(2-tailed)	0.058	0.000	0.000	

* Correlation is significant at the 0.05 level (2-tailed).

** Correlation is significant at the 0.01 level (2-tailed).

<Table 14> Model Summary (2014~2018)

Model	R	R ²	Sig. F Change	Durbin-Watson
1	0.950 ^a	0.902	0.000	
2	0.977 ^b	0.954	0.007	2.223

^a. Predictors: (Constant), Gender equality problems(2014~2018).

^b. Predictors: (Constant), Gender equality problems(2014~2018), Political Issue(2014~2018).

^c. Dependent Variable: Social Problems(2014~2018).

<Table 15> Coefficients^a(2014~2018)

Model		Unstandardized B	Coefficients Std.Error	Standardized Coefficients Beta	t	Sig.	VIF
1	Constant	1155.344	577.473		2.001	0.071	
	Gender conflict(2014_2018)	1.512	0.150	0.950	10.052	0.000	1.000
2	Constant	976.513	419.038		2.330	0.042	
	Gender conflict(2014_2018)	1.381	0.115	0.867	11.893	0.000	1.132
	Political issue (2014_2018)	0.083	0.025	0.242	3.352	0.007	1.132

^a. Dependent Variable: Social Problems(2014~2018).

로 분석 되었다.

5. 결론

본 연구에서는 10년간 발생한 인터넷 뉴스 빅 데이터를 수집하고, 이를 증가율과 빈도수를 기준으로 뉴스 발생 기간에 따른 핵심키워드와 그 변화를 분석한 연구이다. 이를 위해 Keyword Issue Framework를 연구하고, 이를 활용하여, 우리나라에서 사회적 사건에 대한 이슈가 지속적으로 발생되고 있음을 분석하였다. 또한, 10년 동안의 뉴스키워드들의 흐름을 분석하면, 2사분면에 있던 다양한 키워드들의 증가율과, 빈도수가 감소한 반면 4사분면에 있던 키워드들이 1사분면이나 3사분면으로 이동한 것을 알 수 있다. 이를 통해 과거에 비해 최근 더욱 관심이 되고 있는 이슈들을 분석할 수 있었다. 아울러, 본 연구에서는 키워드들의 요인분석을 통해 ‘사회적 사건’, ‘교육문화’, ‘젠더갈등’ 그리고 ‘정치적 이슈’라는 4가지의 잠재변수를 도출하였다. 그리고, 각각의 상관관계 분석을 통해 변수들 간의 방향성을 분석하였다. 이를 바탕으로 다중회귀 분석을 통한 인과관계를 분석하였으며 분석결과 뉴스의 발생 기간에 따라, ‘사회적 사건’에 영향을 미치는 요인들은 각각 상이함을 알 수 있었다.

그러나 ‘젠더갈등(여성, 논란 등)’ 이슈의 발생빈도와 증가율은 시간이 흐름에 따라 상대적으로 감소했음에도 불구하고, ‘사회적 사건’에 지속적으로 영향을 미치는 것으로 분석되었다. 또한 최근(2014~2018) 들어, ‘정치적 이슈’가 사회적 사건에 영향을 미치는 잠재요인으로 분석되었다.

따라서, ‘젠더갈등’, ‘정치적 이슈’같은 갈등적 요인이 최근의 우리나라의 ‘사회적 사건’에 영향을 미치는 주요 요인임으로 분석되었다.

이것은, 최근의 우리나라가 과거와는 다른 새로운 갈등으로 사회문제가 심화될 수 있는 구조로 변화되고 있음을 알 수 있다.

6. 연구의 한계점

본 연구는 언론사의 인터넷 뉴스(사회면)를 통해 발표된 자료만을 수집하여 분석된 결과이다. 따라서, 시민들이 개인 소셜 네트워크 서비스(SNS)를 통해 개인적인 의견을 표출하거나 우리나라가 관심 있게 논의하고 있는 키워드와는 차이가 있을 수 있다. 특히 최근 들어 다양한 매체가 등장하면서 편향되고 거짓된 뉴스(fake news)가 등장하고 있다. 이러한 현상들은 데이터 수집과 분석에 오류를 야기할 수 있다. 따라서 이에 대한 심도 깊은 논의가 요구된다. 또한 불연속적인 데이터의 변화에 대한 흐름을 분석할 수 있는 추가적인 연구가 요구된다.

References

- [1] Ansoff, I. H., “Strategic response in turbulent environments,” Working Paper No. 82-35, European Institute for Advanced Studies in Management, August, 1982.
- [2] Ansoff, I. H., “Managing strategic surprise by response to weak signals,” California Management Review, Vol. XVIII, No. 2,

- pp. 21-33, 1975.
- [3] BIS, Horizon Scanning Centre, 2011. Retrieved from <http://www.bis.gov.uk/foresight/our-work/horizon-scanning-centre>.
- [4] Coffman, B., "Weak Signal Research, Part I: Introduction," 1997.
- [5] Coffman, B., "Weak Signal Research, Part II: Information Theory," 1997.
- [6] Coffman, B., "Weak Signal Research, Part III: Sampling, Uncertainty and Phase Shifts in Weak Signal Evolution," 1997.
- [7] Coffman, B., "Weak Signal Research, Part IV: Evolution and Growth of the Weak Signal to Maturity," 1997.
- [8] Coffman, B., "Part V: A Process Model for Weak Signal Research," 1997.
- [9] Day, G. and Shoemaker, P., "Scanning the Periphery," *Harvard Business Review*, November, pp. 135-148, 2005.
- [10] Downs, A., "Up and Down with Ecology-The Issue-Attention Cycle," *International Library of Comparative Public Policy*, Vol. 11, No. 1, pp. 145-157, 1999.
- [11] Dunlap, R. E. and Scarce, R., "Poll trends: Environmental problems and protection," *The public opinion quarterly*, Vol. 55, No. 4, pp. 651-672, 1991.
- [12] Hasan, N. and Simmons, J. R., "The issue-attention recycle: More ups than downs anticipated," In *A Paper Presented at the Midwest Political Science Association Meeting at Chicago, Illinois*, 1989.
- [13] Henderson, B. D., "Henderson on corporate strategy," *Signet*, 1982.
- [14] Hiltunen, E., "The future sign and its three dimensions," *Futures*, Vol. 40, pp. 247-260, 2008.
- [15] Ilmola, L. and Kuusi, O., "Filters of weak signals hinder foresight: Monitoring weak signals efficiently in corporate decision-making," *Futures*, Vol. 38, No. 8, pp. 908-924, 2006.
- [16] Jeong, G., "A study of future prediction method using text mining and network analysis," Seoul: Korea Institute of Science & Technology Evaluation and Planning, 2014.
- [17] Lee, H. J., "A Study on Process for Draw out Solutions to Social Issues," *Journal of the Korea Management Engineers Society*, Vol. 19 No. 2, pp. 117-127, 2014.
- [18] Lee, D. J., Won, J. U., Kwon, Y. J., and Kim, M. R., "A Study on Customer Satisfaction for Courier Companies based on SNS Big data," *The Journal of Society for e-Business Studies* Vol. 21, No. 4, pp. 55-67, 2016.
- [19] Moijanen, M., "Heikot signaalit tulevaisuuden tutkimuksessa (Weak signals in futures studies)," *Futura*, Vol. 4, pp. 38-60, 2003.
- [20] Park, C. and Kim, H., "A study of development direction of new industries through the internet of things-detecting future signals using text mining," Korea Energy Economics Institute, 2015.
- [21] Pirinen, O., "Weak Signal based foresight

- service,” 2010.
- [22] Rossel, P., “Weak signals as a flexible framing space for enhanced management and decision-making,” In *Foresight for Dynamic Organisations in Unstable Environments*, pp. 31-44, 2013.
- [23] Salton, G. and Buckley, C., “Term-weighting approaches in automatic text retrieval,” *Information Processing & Management*, Vol. 24, No. 5, 513-523, 1988(a, b).
- [24] SBI, *The Scan Process*, 2011, Retrieved from <http://www.strategicbusinessinsig>
[hts.comscan/process.shtml](http://www.strategicbusinessinsig).
- [25] TrendWiki, 2011, Retrieved from <http://www.trendwiki.fi/en>.
- [26] Yoo, S.-H., Park, H.-W., and Kim, K.-H., “A study on exploring weak signals of technology innovation using informatics,” *Journal of Technology Innovation*, Vol. 17, No. 2, pp. 109-130, 2009.
- [27] Yoon, J., “Detecting weak signals for long-term business opportunities using text mining of Web news,” *Expert Systems with Applications*, Vol. 39, No. 16, pp. 12543-12550.

저 자 소 개



이홍주

2010년~현재

2014년~2016년

관심분야

(E-mail: blue1024@kgu.ac.kr)

경기대학교 산업경영공학과 부교수

경기대학교 창업교육센터장

비즈니스 인텔리전스, 비즈니스 모델링, 소비자 행동 분석