

A Comparative Analysis of Cognitive Change about Big Data Using Social Media Data Analysis

Youdong Yun[†] · Jaechoon Jo^{††} · Yuna Hur[†] · Heuseok Lim^{†††}

ABSTRACT

Recently, with the spread of smart device and the introduction of web services, the data is rapidly increasing online, and it is utilized in various fields. In particular, the emergence of social media in the big data field has led to a rapid increase in the amount of unstructured data. In order to extract meaningful information from such unstructured data, interest in big data technology has increased in various fields. Big data is becoming a key resource in many areas. Big data's prospects for the future are positive, but concerns about data breaches and privacy are constantly being addressed. On this subject of big data, where positive and negative views coexist, the research of analyzing people's opinions currently lack. In this study, we compared the changes in peoples perception on big data based on unstructured data collected from the social media using a text mining. As a results, yearly keywords for domestic big data, declining positive opinions, and increasing negative opinions were observed. Based on these results, we could predict the flow of domestic big data.

Keywords : Social Media, Big Data, Opinion Mining, Word Cloud, Annual Key-Words

소셜 미디어 데이터 분석을 활용한 빅데이터에 대한 인식 변화 비교 분석

윤 유 동[†] · 조 재 춘^{††} · 허 윤 아[†] · 임 희 석^{†††}

요 약

최근 모바일의 확산과 웹 서비스의 도입으로 온라인 상에 데이터가 급격히 증가하게 되어 다양한 분야에서 활용되고 있다. 특히, 빅데이터 분야에서 소셜 미디어의 등장은 축적되는 비정형 데이터의 양이 급격하게 증가하는 계기가 되었다. 이러한 비정형 데이터로부터 의미 있는 정보를 추출하기 위해 다양한 분야에서 빅데이터 기술에 대한 관심이 증가하고 있다. 빅데이터는 선진국을 중심으로 다양한 분야에서 핵심 자원으로 중요성이 부각되고 있다. 그러나 빅데이터의 긍정적인 미래 전망과 함께 데이터의 침해 및 개인정보 보호에 대한 우려가 지속적으로 언급되고 있다. 이와 같이 긍정적인 시각과 부정적인 시각이 공존하는 빅데이터에 대해 사람들의 의견을 분석하는 연구는 현재 매우 부족한 상황이다. 이에 본 연구에서는 텍스트 마이닝을 활용하여 소셜 미디어에서 수집한 비정형 데이터를 기반으로 빅데이터에 대한 사람들의 인식 변화를 비교하였다. 텍스트 마이닝 결과, 국내 빅데이터에 대한 연도별 키워드와 함께 시간의 흐름에 따라 감소하는 긍정적인 의견과 증가하는 부정적인 의견이 관찰되었다. 그리고 이러한 분석 결과를 기반으로 국내 빅데이터에 대한 흐름을 예측할 수 있었다.

키워드 : 소셜 미디어, 빅데이터, 오피니언 마이닝, 워드 클라우드, 연도별 키워드

1. 서 론

최근 모바일의 확산과 웹 서비스의 도입으로 인해 온라인 상에 정형 데이터(Formal Data) 뿐만 아니라 비정형 데이터

(Informal Data)가 급격히 증가하게 되어, 다양한 분야에서 다양한 방법으로 활용되고 있다. 빅데이터의 경우 활용성이 높아 세계시장에서 2014년부터 2019년까지 23.1%의 연평균 성장률이 전망되었으며, 국내시장에서는 2014년부터 2018년까지 26.4%의 연평균 성장률이 전망되었다[1]. 특히, 빅데이터 분야에서 소셜 미디어(Social Media)의 등장은 시간과 장소에 관계없이 개인과 조직으로부터 축적되는 비정형 데이터의 양이 보다 빠르게 확산 및 축적되는 계기가 되었다. 실제로 현재 생성되는 디지털 데이터의 약 70%가 전자우편을 포함하여 사용자가 데이터를 생성하는 각종 소셜 미디어

※ 이 논문은 2017년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. R1610941).

† 준 회 원 : 고려대학교 컴퓨터학과 석사과정

†† 준 회 원 : 고려대학교 컴퓨터학과 박사과정수료

††† 정 회 원 : 고려대학교 컴퓨터학과 교수

Manuscript Received : March 8, 2017

First Revision : April 24, 2017

Accepted : May 3, 2017

* Corresponding Author : Heuseok Lim(limhseok@korea.ac.kr)

어에서 생성되고 있다[2]. 이러한 소셜 미디어에서 생성되는 수많은 비정형 데이터로부터 의미 있는 정보를 추출하기 위해 다양한 분야에서 빅데이터 기술에 대한 관심이 증가하고 있으며, 빅데이터가 어떻게 효율적으로 관리 및 분석될 것인지에 대한 지속적인 논의가 이루어지고 있다[3, 4].

일반적으로 빅데이터(Big Data)는 기존의 데이터베이스 소프트웨어가 저장, 관리, 분석할 수 있는 범위를 넘어선 대규모의 데이터를 의미한다. 그러나 빅데이터를 단순히 규모만으로 정의할 수는 없다[5, 6]. 빅데이터는 다양한 환경에서 생성되는 데이터로 일반적인 데이터와 비교하여 데이터의 크기가 방대하고, 데이터의 생성 속도가 매우 빠르며 정형 데이터 뿐만 아니라 텍스트, 이미지, 영상, 음성 등 다양한 데이터 형태를 포함하는 대규모 데이터를 의미하기 때문이다. 즉, 빅데이터는 데이터의 크기(Volume), 데이터의 생성 속도(Velocity), 정보형태의 다양성(Variety)의 세 가지 측면에서 기존의 데이터와는 다른 특징을 보이고 있다 [7, 8]. 이러한 세 가지 측면을 일반적으로 “3V”라고 부르며, 최근에는 “3V”에 네 번째 측면인 빅데이터의 가치(Value)를 포함하여 “4V”로 정의하기도 한다.

빅데이터의 등장은 인간의 삶에 다양한 변화를 가져왔다. 컴퓨터 및 처리기술의 발전으로 빅데이터의 분석이 가능해졌고, 이에 따라 빅데이터가 선진국을 중심으로 산업, 공공, 의료, 과학 등의 다양한 분야에서 핵심 자원 또는 도구로서의 중요성이 부각되고 있다[9]. 그러나 빅데이터의 긍정적인 미래 전망과 함께 지속적으로 언급되는 문제점 중 하나가 데이터의 침해 및 개인정보 보호와 관련된 것이다. 빅데이터 시대에서 다양한 스마트 기기를 통해 수집 및 생성되는 위치정보, 검색패턴, 접속기록과 같은 그림자 데이터(Digital Shadow)는 물론, 본인의 의지로 생성 및 공개한 데이터의 경우에도 이러한 정보가 의도하지 않은 방향으로 활용되거나 악용되면서 개인정보 침해 관련 문제에 대한 발생 가능성이 지속적으로 증가하고 있다[10-11]. 이와 같은 문제를 안고 특정 분야에서 빅데이터 분석에 대한 연구 및 빅데이터 보안에 대한 연구가 활발하게 이루어지고 있다[12-18]. 그러나 사람들이 빅데이터에 대해 어떻게 생각하고 있는지에 대한 의견을 분석하는 연구는 매우 부족한 실정이다.

이에 본 연구에서는 트위터나 블로그와 같은 소셜 미디어에서 수집한 비정형 데이터를 기반으로 의미 있는 정보를 추출하기 위해 텍스트 마이닝 기법을 활용하여 다음과 같은 분석을 실시한다. 첫째, 연도별 빅데이터와 관련된 키워드를 살펴보고, 각 키워드가 어떤 의미를 가지고 있는지 파악한다. 둘째, 연도의 변화에 따라 빅데이터와 관련된 키워드가 어떻게 변화하는지 파악한다. 셋째, 연도별 빅데이터와 관련한 긍정적, 부정적인 감정을 파악하여 빅데이터에 대한 사람들의 인식을 파악한다.

본 연구에서는 소셜 미디어 데이터 분석 결과를 기반으로 빅데이터에 대한 사람들의 인식과 트렌드(Trend)를 파악하는 것을 목적으로 한다. 구체적으로는 소셜 미디어에서 최근 빅데이터에 대한 사람들의 의견을 수집하고, 텍스트 마

이닝 기법을 기반으로 연도별 분석 결과를 도출하여 연도별 빅데이터에 대한 전반적인 흐름을 예측할 수 있다.

본 논문의 2장에서는 본 연구에서 수행하는 텍스트 마이닝 기법의 선행 연구에 대한 내용을 다룬다. 3장에서는 본 연구에서 사용하는 데이터와 함께, 연구 방법에 대해 설명한다. 4장에서는 텍스트 마이닝 분석 결과에 대해 설명하며, 5장에서는 본 연구의 결론 및 향후 연구에 대해 논의한다.

2. 관련 연구

2.1 소셜 미디어 데이터 분석

최근 디지털 기술의 발달과 인터넷 및 스마트 디바이스의 보급으로 인해 소셜 미디어에 대한 접근성이 향상되었고, 이로 인해 소셜 미디어는 급속한 성장을 맞이하게 되었다. 특히, 수많은 사용자들이 스마트 디바이스를 기반으로 소셜 미디어와 같은 정보 채널을 활용하여 실시간으로 대량의 데이터를 생산하기 때문에, 이를 수집, 분석하는 것이 가능해졌다[19]. 이에 따라 소셜 미디어에서 수집된 데이터를 기반으로 사용자들의 생각을 분석하고, 의미 있는 정보를 추출하거나 특정 분야의 흐름을 예상하는 분석 기술들을 활용한 연구가 다양한 분야에서 활발하게 수행되고 있다[20].

J. S. Hong(2016)는 소셜 미디어에서 특정 기업에 대한 데이터를 기업의 위기 전후로 나누어 이미지를 평가하는 연구를 수행하였다[21]. Y. S. Choi(2016)는 소셜 미디어가 외교활동에 있어 영향력을 가지고 효과적인 외교의 소통 수단이 될 수 있는지를 검증하고자 연구를 수행하였다[22]. I. K. Oh(2015)는 국내 관광환경에 대한 관광객들의 관광인식을 소셜 미디어의 빅데이터를 통해 분석하는 연구를 수행하였다[23]. A. R. Lee(2015)는 특정 소셜 미디어를 대상으로 광고에 대한 데이터를 수집하여 실시간으로 광고 효과를 분석하는 시스템을 설계 및 구축하였다[24]. D. Y. Kim(2014)는 주가 예측의 정확도를 높이기 위해 소셜 미디어의 데이터와 뉴스기사 데이터를 동시에 이용한 다수의 주가예측 모형을 생성한 후 정확성을 비교하는 연구를 진행하였다[25].

2.2 소셜 미디어 데이터 감성 분석

텍스트 마이닝(Text Mining)의 한 분류인 오피니언 마이닝(Opinion Mining)은 정형, 비정형 텍스트 데이터에서 긍정, 부정 또는 중립의 의견을 판단하고 활용하는 목적으로 사용되며, 감성 분석(Sentiment Analysis)이라고 하기도 한다. 소셜 미디어에서 수집되는 데이터는 주관적인 의견의 비율이 높기 때문에 기존의 웹 문서에 비해 오피니언 마이닝을 적용하기에 좋은 환경을 갖추었다고 할 수 있다. 실제로 실시간으로 대량의 의견 데이터를 생산하는 소셜 미디어와 같은 정보 채널로 인해 의견 데이터에 대한 수집이 원활해졌고, 이를 수집하여 사용자의 의견을 분석하고자 하는 오피니언 마이닝이 각광받고 있다[26-28].

K. S. Song(2015)는 소셜 미디어의 빅데이터를 활용하여 개인의 취향과 감성에 따라 영화, 연극, 음악 등의 문화 콘

텐츠를 추천해 주는 서비스 모형을 제안하였다[29]. S. N. Kang(2015)는 소셜 미디어에서 작성되는 사회 이슈에 대한 의견 데이터를 수집하여 감성사전을 통한 감성 분석을 수행하였다[30]. Y. M. Kim(2014)는 소셜 미디어 상의 데이터를 수집하여 주식시장 참여자들의 감성을 분석하고 이를 기반으로 주가 등락을 예측하는 절차를 제안하였다[28]. Y. J. Lee(2014)는 소셜 미디어 텍스트 데이터를 활용하여 연관 규칙 기반의 오피니언 감성사전을 구축하고, 사용자의 의견을 분석하여 의류 마케팅에 활용할 수 있는 감성 트렌드 예측 분석 알고리즘을 제안했다[31].

3. 연구 방법

본 연구는 소셜 미디어를 통해 빅데이터에 대한 사람들의 의견 데이터를 수집하고, 텍스트 마이닝 기법을 활용하여 연도별 키워드(Key-words)를 통해 트렌드를 확인한다. 그리고 연도별로 데이터를 비교, 분석하여 빅데이터에 대해 변화하는 사람들의 인식을 파악할 수 있도록 한다. 데이터에 대한 분석 과정은 Fig. 1과 같이 크게 데이터 수집, 데이터 전처리, 텍스트 마이닝의 세 가지 단계로 진행되었다.

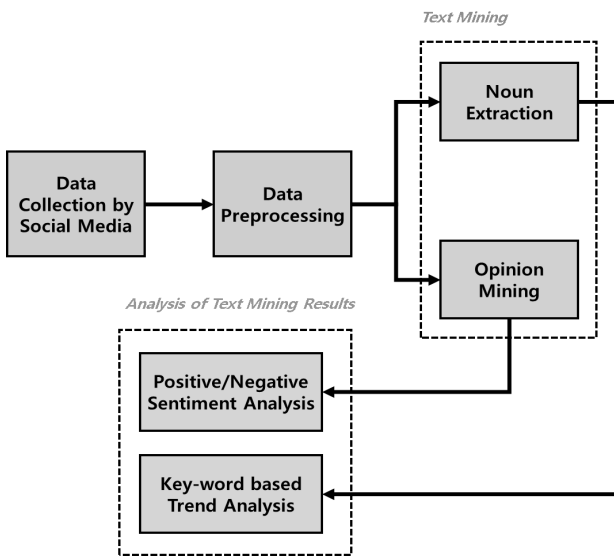


Fig. 1. Courses of Research

3.1 소셜 미디어 데이터 수집

본 연구에서 사용하는 데이터는 트위터나 블로그와 같은 소셜 미디어에 사람들이 작성한 빅데이터와 관련된 의견 데이터를 2014, 2015, 2016, 2017년도로 분류하여 수집하였으며, 2017년도 데이터는 3월 이전의 데이터만을 수집하였다. 2017년도 데이터는 전년도에 비해 짧은 기간 동안 수집되어 비교적 적은 데이터가 수집되었으며, 이에 대한 자세한 내용은 다음 Table 1에 나타나 있다.

본 연구에서 활용한 의견 데이터는 2014년도에 총 83,189 (25.46%)개의 의견 데이터를 수집하였으며, 2015년도에는 총

78,322(23.97%)개의 의견 데이터를 수집하였다. 그리고 2016년도에는 총 138,082(42.25%)개의 의견 데이터를 수집하였으며, 2017년도에는 총 27,193(8.32%)개의 의견 데이터를 수집하여 총 326,786(100%)개의 의견 데이터로 구성되어 있다.

Table 1. Result of Data Collection

Year	Number of Posts(Twitter)	Number of Posts(Blog)	Total	Rate(%)
2014	63,365	19,824	83,189	25.46
2015	52,620	25,702	78,322	23.97
2016	106,470	31,612	138,082	42.25
2017	21,821	5,372	27,193	8.32
Total	244,276	82,510	326,786	100

3.2 데이터 전처리

수집된 소셜 미디어 의견 데이터는 일반적으로 텍스트 데이터로 이루어져 있다. 따라서 데이터 분석을 위해 우선적으로 분석이 가능한 형태로 수집된 데이터를 정제하는 과정이 필요하다. 전처리 작업은 다양한 종류의 비정형 데이터를 분석할 수 있도록 해주는 처리 단계라고 할 수 있다. 이러한 과정을 거치지 않으면 필요한 속성이 없는 불완전(Incomplete) 문제 및 의미가 맞지 않는 값이 존재하는 노이즈(Noisy) 문제 등이 나타날 수 있다. 이에 따라 수집한 소셜 미디어 의견 데이터에 대해 전처리 작업으로 @트윗, URL, 문장 부호, 숫자, 불용어에 대한 제거를 수행하였다[32-34].

3.3 소셜 미디어 데이터 분석

본 연구에서는 수집된 소셜 미디어 의견 데이터를 기반으로 명사 추출(Noun Extraction), 오피니언 마이닝(Opinion Mining)의 두 가지 분석을 수행한다. 명사 추출을 통해 각 연도별로 언급된 키워드를 비교 및 분석하여 빅데이터에 대해 어떤 트렌드 변화가 있는지 파악한다. 그리고 오피니언 마이닝을 통해 사람들이 빅데이터에 대해 어떤 생각을 가지고 있는지 판단하고, 이에 대한 변화를 살펴보고자 한다.

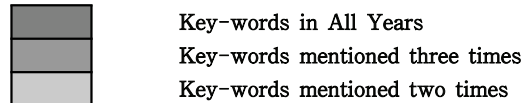
4. 연구 결과

4.1 명사 추출

다음 Table 1은 빅데이터에 대한 사람들의 의견 데이터에서 등장한 빈도 수를 기반으로 특정 인물 키워드를 제외하고 상위 15개씩 나타낸 결과이다. 명사 추출 결과 2014, 2015, 2016, 2017년도에 공통적으로 상위 15개 키워드에 나타난 단어로는 ‘분석’, ‘기술’, ‘정보’, ‘서비스’의 네 단어로 나타났다. 2014, 2015, 2016년도에 공통적으로 상위 15개 키워드에 나타난 단어로는 ‘기업’이 나타났다. 그리고 2014, 2015년도에 공통적으로 상위 15개 키워드에 나타난 단어로는 ‘인터넷’, ‘시스템’, 2016, 2017년도에 공통적으로 상위 15개 키워드에 나타난 단어로는 ‘지능’이 나타났다.

Table 2. Results of Noun Extraction

No.	2014	Freq.	2015	Freq.	2016	Freq.	2017	Freq.
1	Blog	16,931	Analysis	13,102	Analysis	26,654	Analysis	5,811
2	Technology	14,092	Technology	11,330	Technology	12,617	Presidential Election	3,515
3	Information	13,334	Information	10,601	Information	12,070	Result	2,491
4	Analysis	11,250	Service	6,582	Service	11,192	Public Opinion	2,358
5	Service	7,425	Enterprise	6,303	Result	10,040	Technology	2,179
6	Enterprise	6,348	Internet	5,631	Nation	7,482	Public Opinion Poll	1,994
7	Naver	4,869	Expert	5,038	Seoul	6,926	Reputation	1,938
8	Personal Information	4,862	System	4,758	Seoul City	6,713	Service	1,926
9	Database	4,597	Industry	4,688	Enterprise	6,555	Information	1,812
10	Internet	4,588	Result	4,478	Subway	6,514	Google	1,730
11	System	4,137	Market	4,204	Participant	6,369	Brand	1,656
12	Resource	4,126	Base	4,175	Numeration	6,358	Power	1,463
13	Open Source	4,066	Idol	3,867	Passenger	6,260	News	1,416
14	Specialty	4,050	Business	3,474	Police	6,239	Intelligence	1,380
15	Source	3,864	Development	3,343	Intelligence	6,066	Artificial Intelligence	1,288



최근 빅데이터가 등장하고, 다양한 분야에서 빅데이터 기술에 대한 관심이 증가함에 따라 빅데이터가 어떻게 효율적으로 관리 및 분석될 것인지에 대한 논의가 지속되고 있다. 이러한 사회의 흐름에 따라 ‘분석’, ‘기술’, ‘정보’가 연도에 관계없이 공통적으로 등장하였으며, 빅데이터 기술에 대한 서비스가 지속적으로 등장하면서 ‘서비스’와 같은 키워드가 나타났다. 그리고 2015, 2016, 2017년도에는 빅데이터 분석에 대한 관심이 증가함에 따라 ‘결과’라는 키워드가 공통적으로 나타났다. 2010년 이후에는 빅데이터가 새로운 기업가치 창출의 핵심요인으로 전망된 이후 비즈니스 분석(Business Analytics)분야는 데이터를 통해 기업에서 발생하는 문제에 대한 해결 방안으로 활용되고 있다[35]. 이러한 흐름에 따라 2014, 2015, 2016년도에는 ‘기업’이라는 키워드가 나타났다. 2014, 2015년도에는 빅데이터에 대한 관심이 증가함에 따라 빅데이터가 수집되는 인터넷 환경 및 빅데이터 분석을 수행할 수 있는 시스템에 대한 관심도 함께 증가하여 ‘인터넷’, ‘시스템’과 같은 키워드가 나타났다. 인공지능에 대한 연구는 오랜 시간 이루어졌으나 최근 빅데이터 기술의 등장은 인공지능 분야에 적지 않은 영향을 미쳤다. 빅데이터 시대에는

데이터의 종류와 양이 증가하고 분석 능력이 향상됨에 따라 데이터와 분석 기술이 기반이 되는 인공지능의 구현이 가속화 될 수 있었다[1]. 이에 따라 2016년도 이후 ‘지능’, ‘인공지능’과 같은 키워드가 새롭게 주목받고 있다. 그리고 연도별 명사 추출 결과를 보면, 연도별로 화두가 되는 사건에 따라 빅데이터에 대한 키워드가 결정되는 모습을 보이기도 했다. 그 예로 2016년도의 ‘지하철’, ‘승객’, 2017년도의 ‘대선’, ‘여론’, ‘여론조사’와 같이 특정 연도에 화두가 된 사건을 중심으로 각 연도의 독립적인 키워드를 확인할 수 있다.

4.2 워드 클라우드

워드 클라우드(Word Cloud)는 문서의 키워드나 개념을 직관적으로 파악할 수 있도록 핵심 단어를 시각적으로 돋보이게 하는 기법이다. 본 연구에서는 워드 클라우드를 활용하여 명사 추출을 통해 추출한 의견 데이터의 키워드를 직관적으로 파악할 수 있도록 빅데이터 의견 데이터에서 나타난 키워드의 빈도 수에 따라 단어의 크기를 다르게 하였다. 다음 Fig. 2에서 명사 추출 결과 나타난 키워드를 워드 클라우드를 통해 직관적으로 파악할 수 있다.

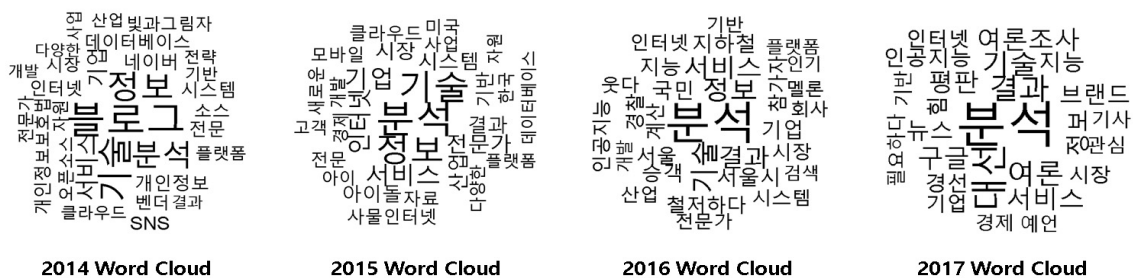


Fig. 2. Word Cloud Based on Noun Extraction Result

4.3 오피니언 마이닝

본 연구에서 수행하는 오피니언 마이닝은 소셜 미디어에서 수집한 빅데이터에 대한 의견 데이터가 갖는 긍정, 부정 또는 중립적인 의견 중에서 어떠한 의견을 가지고 있는지 판별하는 과정으로서, 문서 최소 단위인 어휘의 감성 극성을 기반으로 이루어진다[36, 37]. 오피니언 마이닝은 미리 구축된 감성 사전과, 소셜 미디어 상에 작성된 의견을 매칭하여 텍스트 내 극성을 분석하게 된다. 이러한 분석 과정으로 오피니언 마이닝의 결과는 감성 사전(Sentiment Dictionary)의 품질에 따라 다르게 나타날 수 있다[38].

본 연구에서 사용한 감성 사전은 총 882개의 긍정 단어와 1,679개의 부정 단어로 구성되어 있다. 다음 Table 3에서 긍정 단어와 부정 단어에 대한 예시가 나타나 있다.

Table 3. Hangeul Positive, Negative Words Example

Year	Words Example
Positive Words	가능, 가치, 간결, 감사, 감탄, 낭만, 노력, 낙관, 다정, 도움, 독창, 동기, 만족, 맞춤, 매력, 명랑, 명성, 반응, 불의, 비상, 상냥, 상쾌, 생각, 선견, 선량, 선의, 선택, 선호, 순수, 안심, 안정, 애정, 이상, 이해, 인기, 자선, 자신감, 재능, 재미, 편리, 편안, 해방, 행복, 행운,
Negative Words	가격, 가난, 가망, 가뭄, 가지, 격노, 기근, 기만, 나태, 낙심, 난색, 낭비, 낭패, 단절, 단순, 답답, 당황, 무능, 무력, 무모, 무식, 미숙, 반대, 반동, 반란, 반항, 배신, 악명, 악몽, 악인, 양심, 암중, 예민, 오류, 절망, 절박, 저해, 좌절, 주저, 죽음, 편협, 포기, 혼란, 현기증,

다음 Fig. 3은 소셜 미디어에서 추출한 2014년도부터 2017년도까지의 빅데이터 의견 데이터에 대한 오피니언 마이닝 결과를 나타낸다. 2014년도 오피니언 마이닝 결과, 많은 빈도를 나타낸 100개 연관어 중 긍정을 나타내는 단어가 40개, 중립을 나타내는 단어가 46개, 부정을 나타내는 단어가 14개로 나타났다. 2015년도 오피니언 마이닝 결과, 많은 빈도를 나타낸 100개 연관어 중 긍정을 나타내는 단어가 47개, 중립을 나타내는 단어가 42개, 부정을 나타내는 단어가 11개로 나타났다. 2016년도 오피니언 마이닝 결과, 많은 빈도를 나타낸 100개 연관어 중 긍정을 나타내는 단어가 47개, 중립을 나타내는 단어가 34개, 부정을 나타내는 단어가 19개로 나타났다. 2017년도 오피니언 마이닝 결과, 많은 빈도를 나타낸 100개 연관어 중 긍정을 나타내는 단어가 47개, 중립을 나타내는 단어가 34개, 부정을 나타내는 단어가 19개로 2014년도부터 2017년도까지 모든 연도에서 빅데이터에 대해 대체적으로 긍정적인 반응을 보이는 것으로 나타났다.

분석 결과에서 확인할 수 있듯이, 소셜 미디어 상에서 사람들은 아직 빅데이터에 대해 긍정적인 반응을 보이는 것으로 나타났다. 이러한 결과는 최근 빅데이터 기술의 발달로 인해 다양한 분야에서 빅데이터를 활용하여 새로운 발견 및 융합뿐만 아니라 문제 해결을 위한 핵심적인 도구로서 활용되는 시점에서 발생한 의견을 분석하여 도출된 결과이다. 그러나 2014, 2015년도에 비해 2016년 이후 부정적인 의견이

증가하였다. 이러한 결과를 통해 앞으로 빅데이터에 대해 더 많은 시간을 지켜볼 필요가 있다고 판단된다.

다음 [Fig 4]은 2014, 2015, 2016, 2017년도 빅데이터에 대한 오피니언 마이닝 결과를 선형 그래프를 통해 나타낸 것이다. 그래프에서 긍정적인 의견은 2014년도에 비해 2015, 2016년도에 증가하였으나, 2016년도 이후 감소하는 추세를 보이고 있다. 반면, 부정적인 의견은 2014, 2015년도에 비해 2016, 2017년도에 증가한 모습을 확인할 수 있다.

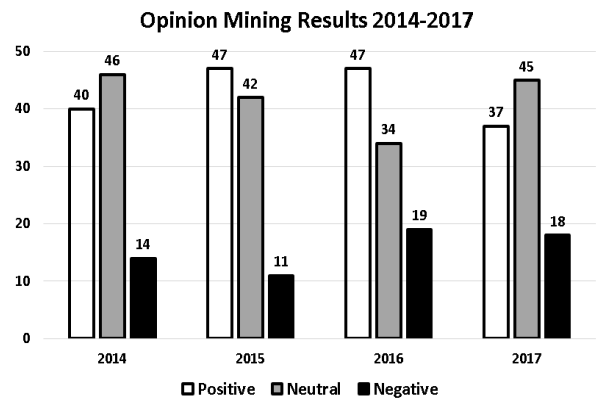


Fig. 3. 2014-2017 Opinion Mining Results

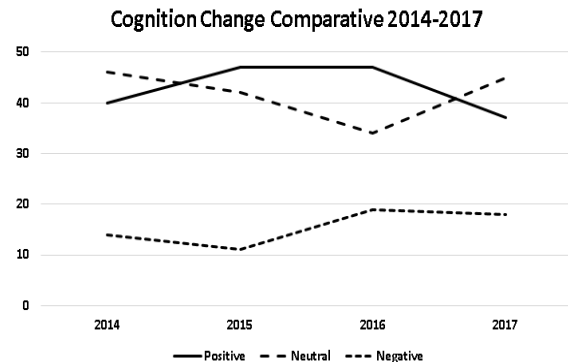


Fig. 4. 2014-2017 Cognition Change Comparative

5. 결론 및 제언

최근 컴퓨터 및 처리기술의 발전과 함께 빅데이터의 기술도 함께 발전하여 빅데이터가 선진국을 중심으로 산업, 공공, 의료, 과학 등의 다양한 분야에서 핵심 자원 또는 도구로서 주목받고 있다. 그러나 빅데이터에 대한 긍정적인 시각과 함께 데이터의 침해 및 개인정보 보호와 같이 부정적인 시각도 존재한다. 이와 같은 문제를 안고 빅데이터와 관련된 연구가 활발하게 이루어지고 있으나 사람들이 빅데이터에 대해 어떻게 생각하고 있는지에 대한 의견을 분석하는 연구는 매우 부족한 실정이다. 이에 따라 본 연구에서는 소셜 미디어에서 2014년부터 2017년까지 연도별로 수집한 빅데이터 관련 의견 데이터를 기반으로 명사 추출과 오피니언

마이닝을 수행하여 빅데이터에 대한 사람들의 인식이 어떻게 변화하는지 살펴보았다.

명사 추출 결과 모든 연도에 공통적으로 상위 15개 키워드에 나타난 단어로는 ‘분석’, ‘기술’, ‘정보’, ‘서비스’가 나타났다. 2014, 2015, 2016년도에 공통적으로 상위 15개 키워드에 나타난 단어로는 ‘기업’이 나타났다. 그리고 2014, 2015년도에 공통적으로 상위 15개 키워드에 나타난 단어로는 ‘인터넷’, ‘시스템’이, 2016, 2017년도에는 ‘지능’이 나타났다. 이러한 결과를 기반으로 빅데이터가 어떻게 효율적으로 관리 및 분석될 것인지에 대한 관심이 계속될 것으로 보여지며, 특히 2016년 이후로 빅데이터 기술의 발달로 인해 데이터와 분석 기술이 기반이 되는 인공지능에 대한 관심이 지속적인 증가할 것으로 전망된다.

오피니언 마이닝 결과, 소셜 미디어에서 추출한 빅데이터에 대한 사람들의 의견은 전체적으로 긍정적인 의견이 대다수인 것으로 나타났다. 그러나 긍정적인 의견이 2016년도 이후 감소하는 추세를 보이는 것으로 나타났으며, 부정적인 의견은 2014, 2015년도에 비해 증가하는 추세를 보이는 것으로 나타났다. 이러한 결과는 빅데이터 기술이 발달하고, 다양한 분야에서 활용됨에 따라 사람들이 체감하는 데이터의 침해, 개인정보 보호와 같은 문제점에 대한 우려가 커지고 있기 때문일 가능성이 높다. 실제로 빅데이터의 문제점으로 항상 사생활 침해와 보안 측면이 거론되고 있으며[39], 수집된 데이터가 보안 문제로 유출된다면 거의 모든 사람들의 정보가 유출되어 사회적으로 큰 문제를 일으킬 가능성이 있다[40]. 그러나 이러한 문제는 빅데이터 기술의 발달과 함께 항상 언급되어 왔으며, 이로 인해 빅데이터 보안에 대한 연구가 활발하게 이루어지고 있다. 이미 빅데이터는 다양한 분야에서 중요한 역할을 수행하고 있으나, 빅데이터 보안에 대한 적절한 해결책이 마련된다면 빅데이터 기술은 핵심적인 자원 또는 도구로서 무궁무진한 잠재력을 지니고 있다.

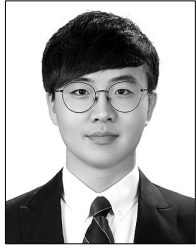
본 연구에서는 소셜 미디어에서 수집한 빅데이터에 대한 사람들의 의견을 수집하고, 명사 추출 방법을 활용하여 연도별로 키워드의 변화 및 키워드의 의미를 파악하였다. 또한 오피니언 마이닝을 활용하여 연도별 빅데이터와 관련한 사람들의 감정을 분석하여 빅데이터에 대한 사람들의 인식을 파악하였다. 이를 통해 국내 빅데이터에 대한 흐름뿐만 아니라, 빅데이터의 미래 전망에 대해 예측해낼 수 있었다.

본 논문은 빅데이터에 대한 사람들의 의견을 분석하기 위해 소셜 미디어에 존재하는 빅데이터에 대한 사람들의 의견 데이터를 수집하였으나, 다음과 같은 한계점이 존재한다. 첫째로, 충분히 검증되고 좋은 품질의 한글 감성 사전의 부재로 오피니언 마이닝 결과에 대한 설득력이 비교적 낮을 수 있다. 둘째로, 의견 데이터의 수집 범위가 소셜 미디어로 제한되어 다양한 의견 데이터에 대한 수집이 이루어지지 않았다. 향후 연구에서는 이와 같은 한계점을 보완하여 연구를 수행한다면 지금보다 더 신뢰도 있고 높은 품질의 분석 결과를 도출해 낼 수 있을 것으로 기대한다.

References

- [1] S. M. Rue, "BigData Effects on Artificial Intelligence," *Journal of Korean Institute of Information Technology*, Vol.14, No.1, pp.29-34, 2016.
- [2] Y. J. Jang and S. K. Cho, "A Comparative Analysis of Data Gathering and Sampling Methods for Social Data," *Social Science Studies*, Vol.25, No.2, pp.3-25, 2014.
- [3] S. H. Yun, K. H. Lee, H. S. Lim, D. R. Kim, and J. H. Kim, "The Method of Digital Copyright Authentication for Contents of Collective Intelligence," *Journal of the Korea Convergence Society*, Vol.6, No.6, pp.185-193, 2015.
- [4] J. H. Kim, J. Y. Go, and K. H. Lee, "A Scheme of Social Engineering Attacks and Countermeasures Using Big Data based Conversion Voice Phishing," *Journal of the Korea Convergence Society*, Vol.6, No.1, pp.85-91, 2015.
- [5] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A. H. Byers, "Big data: The next frontier for innovation, competition, and productivity," 2011.
- [6] G. H. Han and S. H. Jin, "Introduction to Big Data and the Case Study of Its Applications," *Journal of The Korean Data Analysis Society*, Vol.16, No.3, pp.1337-1351, 2014.
- [7] P. Carter, "Big Data Analytics: Future Architectures, Skills and Roadmaps for the CIO," *White paper*, IDC sponsored by SAS, pp.1-16, 2011.
- [8] A. McAfee, E. Brynjolfsson, T. H. Davenport, D. J. Patil, and D. Barton, "Big data," *The management revolution. Harvard Bus Rev*, Vol.90, No.10, pp.61-67, 2012.
- [9] K. S. Noh and J. Y. Lee, "A Study on Analysis of the Differences for Perception of Big Data in Era of Convergence," *Journal of Digital Convergence*, Vol.13, No.10, pp.305-312, 2015.
- [10] C. W. Park, J. W. Kim, and H. J. Kwon, "An Empirical Research on Information Privacy Risks and Policy Model in the Big data Era," *The Journal of Society for e-Business Studies*, Vol.21, No.1, pp.131-145, 2016.
- [11] H. S. Lee, D. W. Lim, and H. J. Zo, "Personal Information Overload and User Resistance in the Big Data Age," *The Journal of Intelligence and Information Systems*, Vol.19, No.1, pp.125-139, 2013.
- [12] B. C. Kim, "Big Data Security Technology and Response Study," *Journal of Digital Convergence*, Vol.11, No.10, pp.445-451, 2013.
- [13] D. W. Kim, S. J. Baek, and J. S. Lee, "Development of Methodology for Future Disaster Foresight Using Bigdata and Collective Intelligence," *Journal of the Korean Society of Civil Engineers*, Vol.65, No.1, pp.12-18, 2017.
- [14] E. J. Jung and U. J. Chang, "Tendency and Network Analysis of Diet Using Big Data," *Journal of the Korean Dietetic Association*, Vol.22, No.4, pp.310-319, 2016.

- [15] J. Y. Bae and M. S. Cho, "Ontology of Food Safety for the Social Big Data Analysis," *Journal of The Korea Society of Health Informatics and Statistics*, Vol.41, No.3, pp.278-289, 2016.
- [16] Y. T. Kim, M. S. Seen, and S. K. Oh, "The Analysis in the Lifestyles of MICE Participants using Big Data," *Journal of Tourism and Leisure Research*, Vol.28, No.10, pp.21-34, 2016.
- [17] D. J. Lee, J. U. Won, Y. J. Kwon, and M. R. Kim, "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.
- [18] R. U. Park, and K. H. Lee, "Study of the Activation Plan for Rural Tourism of the Jeollabuk-do Using Big Data Analysis," *The Korean Society of Community Living Science*, Vol.27, No.5, pp.665-679, 2016.
- [19] J. Y. Seo and C. Koh, "Big Data Analysis by Sensitivity Analysis," *Journal of the Society of Convergence Knowledge*, Vol.2, No.1, pp.15-21, 2014.
- [20] J. Y. Chang, "Convergence of Education and Information & Communication Technology: A Study on the Communication Characteristics of SNS Affecting Relationship Development between Professor and Student," *Journal of the Korea Convergence Society*, Vol.6, No.6, pp.213-219, 2015.
- [21] J. S. Hong and I. K. Oh, "Image difference of before and after an incident using social big data analysis: Focusing on a ramp return of 'K' airline," *International Journal of Tourism and Hospitality Research*, Vol.30, No.6, pp.119-133, 2016.
- [22] Y. S. Choi and H. M. Kim, "The Influence of Public Diplomacy with Social Media on Country Image and Country Brands -Focusing on Cultural Contents," *The Journal of the Korea Contents Association*, Vol.16, No.3, pp.426-438, 2016.
- [23] I. K. Oh, T. S. Lee, and C. N. Chon, "A Study on Awareness of Korea Tourism through Big Data Analysis," *Journal of Tourism Sciences*, Vol.39, No.10, pp.107-126, 2015.
- [24] A. R. Lee, J. S. Bang, and Y. H. Kim, "A Design of a TV Advertisement Effectiveness Analysis System Using SNS Big-data," *KIISE Transactions on Computing Practices, KTCP*, Vol.21, No.9, pp.579-586, 2015.
- [25] D. Y. Kim, J. W. Park, and J. H. Choi, "A Comparative Study between Stock Price Prediction Models Using Sentiment Analysis and Machine Learning Based on SNS and News Articles," *Journal of the Korea Society of IT services*, Vol.13, No.3, pp.211-233, 2014.
- [26] S. Choi and K. H. Choi, "Achievement and satisfaction research of the undergraduate orchestra club activities - A convergent aspects of statistical method and opinion mining," *Journal of the Korea Convergence Society*, Vol.6, No.4, pp.25-31, 2015.
- [27] Y. M. Kim, S. J. Jeong, and S. J. Lee, "A Study on the Stock Market Prediction Based on Sentiment Analysis of Social Media," *Entrue Journal of Information Technology*, Vol.13, No.3, pp.59-69, 2014.
- [28] J. Y. Jang, "Automatic Retrieval of SNS Opinion Document Using Machine Learning Technique," *The Journal of the Institute of Internet, Broadcasting and Communication*, Vol.13, No.5, pp.27-35, 2013.
- [29] K. S. Song, H. Y. Noh, and S. J. Lee, "Recommendation of emotion-based service by using SNS: A case of movie industry," *Journal of the Korea Management Engineers Society*, Vol.20, No.2, pp.91-104, 2015.
- [30] S. N. Kang, Y. S. Kim, and S. H. Choi, "Study on the social issue sentiment classification using text mining," *Journal of the Korean Data & Information Science Society*, Vol.26, No.5, pp.1167-1173, 2015.
- [31] Y. J. Lee, J. H. Seo, and J. T. Choi, "Fashion Trend Marketing Prediction Analysis Based on Opinion Mining Applying SNS Text Contents," *Journal of Korean Institute of Information Technology*, Vol.12, No.12, pp.163-170, 2014.
- [32] S. H. Jun, "A Big Data Preprocessing using Statistical Text Mining," *Journal of Korean Institute of Intelligent Systems*, Vol.25, No.5, pp.470-476, 2015.
- [33] H. G. Jun, G. S. Hyun, K. B. Lim, W. H. Lee, and H. J. Kim, "Big Data Preprocessing for Predicting Box Office Success," *KIISE Transactions on Computing Practices*, Vol.20, No.12, pp.615-622, 2014.
- [34] J. S. Kim, and S. H. Jin, "A Study on the Application of Opinion Mining Based on Big Data," *Journal of the Korean Data Analysis Society*, Vol.15, No.1, pp.101-113, 2013.
- [35] S. H. Kim, J. S. Park, J. H. Park, and I. H. Kim, "A Study on the Effect of Analytic Resources to Business Performance under Big Data Environments," *The Korean Journal of Bigdata*, Vol.1, No.1, pp.23-32, 2016.
- [36] M. J. Kim and E. J. Song, "A Design of Satisfaction Analysis System For Content Using Opinion Mining of Online Review Data," *Journal of Internet Computing and Services*, Vol.17, No.3, pp.107-113, 2016.
- [37] E. H. Jeong and B. K. Lee, "A Design of SNS Emotional Information Analysis Strategy based on Opinion Mining," *Journal of Korea Institute of Information, Electronics, and Communication Technology*, Vol.8, No.6, pp.544-550, 2015.
- [38] H. Chen and D. Zimbra, "AI and opinion mining," *IEEE Intelligent Systems*, Vol.25, No.3, pp.74-80, 2010.
- [39] C. Tankard, "Big Data Security," *Network Security*, pp.5-8, 2012.
- [40] Y. S. Jeong and K. H. Han, "Service Management Scheme using Security Identification Information adopt to Big Data Environment," *Journal of Digital Convergence*, Vol.11, No.12, pp.393-399, 2013.



윤 유 동

e-mail : 2015010492@korea.ac.kr
2015년 목원대학교 마케팅정보건설링학과
(경제학사)
2015년~현 재 고려대학교 컴퓨터학과
석사과정
관심분야: 데이터마이닝, 빅데이터,
머신러닝, 딥러닝



허 윤 아

e-mail : yj72722@korea.ac.kr
2016년 백석대학교 정보보호학과
(이학학사)
2016년~현 재 고려대학교 컴퓨터학과
석사과정
관심분야: 교육데이터마이닝, 머신러닝



조 재 춘

e-mail : jaechoon@korea.ac.kr
2010년 제주대학교 컴퓨터교육과(이학학사)
2012년 고려대학교 컴퓨터교육과(이학석사)
2012년~현 재 고려대학교 컴퓨터학과
박사과정수료
관심분야: 교육데이터마이닝, 학습 분석,
스마트 교육, SW 교육



임 희 석

e-mail : limhseok@korea.ac.kr
1992년 고려대학교 컴퓨터학과(이학학사)
1994년 고려대학교 컴퓨터학과(이학석사)
1997년 고려대학교 컴퓨터학과(이학박사)
2008년~현 재 고려대학교 컴퓨터학과
교수
관심분야: 자연어처리, 뇌신경 언어 정보 처리