

국내 주요 10대 기업에 대한 국민 감성 분석: 다범주 감성사전을 활용한 빅 데이터 접근법

김서인

한양대학교 경영대학 경영학부
(seoin1991@hanyang.ac.kr)

김동성

한양대학교 일반대학원 경영학과
(paulus82@hanyang.ac.kr)

김종우

한양대학교 경영대학 경영학부
(kjuw@hanyang.ac.kr)

최근에 빅 데이터를 활용하여 감성을 측정하는 시도가 활발히 이루어지고 있다. 통신 매체와 SNS의 발달로 기업은 국민의 감성을 파악하고 즉시 대응해야 할 필요성이 생겼다. 우리나라의 경제는 대기업에 대한 의존도가 높기 때문에 10대 기업에 대한 감성분석은 의미가 있다고 할 수 있다. 이러한 측면에서 본 연구는 다 범주를 기준으로 구축한 감성사전을 활용하여 우리나라 10대 기업에 대한 감성을 분석하였다. 빅 데이터를 이용하여 감성을 분석한 기존의 선행연구는 감성을 차원으로 분류하는 경향이 있다. 차원적 감성으로 감성을 분류하는 것은 분류의 기준이 학술적으로 증명되었기에 감성 분석에 주로 사용되어 왔지만 전문가 정도의 지식이 있어야 분류할 수 있어 보편적인 감성을 대변하는 데 비효과적이기에 보완이 필요하다고 할 수 있다. 개별 범주적 감성은 이 점을 보완할 수 있는 분류 방식으로 일정 수준의 주관성이 개입되지만 보편적으로 느낄 수 있는 감성을 측정하는데 효과적이다. 따라서 본 연구는 보편적인 감성의 측정을 위해 감성을 차원으로 분류하지 않고 개별 범주로 분류하여 9가지 영역으로 나누었다. 선행 연구에서 추출한 9가지 범주에 해당하는 감성 단어를 기초하여 감성사전을 구축하였으며 감성 단어가 검출된 빈도를 기준으로 감성을 분석했다. 대상 데이터는 2014년 1월부터 2016년 1월까지 우리나라 10대 기업에 대하여 축적된 뉴스 데이터이다. 대상 데이터에서 검출된 감성 단어의 빈도를 기준으로 각 기업에 대한 감성 순위를 나누고 분포를 확인하였다. 기업에 따라서 감성이 다를 수 있는지, 특정 사건이 각 기업에 대한 감성에 영향을 줄 수 있는지 가설을 세우고 검증하였다. 결론적으로, 다 범주 감성 사전을 활용한 감성 분석은 기업 간 비교와 시점 간 비교에 유의한 것으로 나타났다. 본 연구는 빅 데이터에 산재해있는 감성을 국민의 시각으로 측정하는 하나의 대안으로서 의미가 있다.

주제어 : 감성분석, 차원적 감성, 범주적 감성, 다 범주 감성사전

논문접수일 : 2016년 8월 18일 논문수정일 : 2016년 9월 13일 게재확정일 : 2016년 9월 24일
원고유형 : 일반논문 교신저자 : 김종우

1. 개요

통신 매체와 소셜 네트워크 서비스의 발달로 인하여 기업과 고객의 관계는 변하고 있으며 이제 기업과 소비자는 시공간을 초월하여 양방향으로 소통할 수 있게 되었다(Kotler, 2010). 동시에, 소비자들 간 정보의 전달 및 공유가 활발해

짐에 따라 소비자는 기업의 경영 활동에 직간접적으로 영향을 미치는 프로슈머(prosumer)가 되기도 한다. 기업은 과거에 품질이나 가격과 같은 상품 자체에 중점을 두었다면, 최근에는 기업 중심의 생각에서 벗어나 고객 중심의 사고를 시도하고 있다. 상품의 가격과 품질은 고도의 산업화로 인해 표준화 되었기에 기업은 고객이 상품을

경험할 때 느끼는 감성에 집중해야 할 필요가 있다고 할 수 있다. 소비자의 감성은 통신 매체나 소셜 네트워크 서비스를 통하여 언제 어디서나 피드백을 할 수 있기에 기업은 온라인에 쌓여 있는 빅 데이터를 이용하여 고객의 감성을 이해하고 즉시 대응해야 할 필요가 있다. 고객은 온라인을 통하여 집단적인 감성을 구축하고 기업에 대한 경험을 공유하며 새로운 정보를 빠르게 확산시킨다. 실제로 상품을 구매하지 않았다고 하더라도 이러한 감성은 집단화되어 다양한 측면에서 기업에 영향을 미친다. 온라인에서 집단화된 국민의 감성은 기업의 브랜드 이미지에 영향을 미칠 수 있으며, 이윤과 같은 기업 성과에도 직접 영향을 줄 수 있다. 그러므로 국민의 감성을 파악하고 대응하는 것은 중요하다고 할 수 있다. 예를 들어, 애플과 같은 기업은 고객의 감성 욕구를 충족시킨 기업으로 평가 받으며 급격히 성장했다. 반면에, 2014년 12월에 일어난 ‘대한항공’의 ‘땅콩회항’ 사건은 국민의 감성에 효과적으로 대응하지 못하여 발생한 부정적인 사례라고 할 수 있다.

최근 기업과 여러 연구기관에서 빅 데이터를 이용하여 고객의 감성을 측정하고, 측정된 데이터를 이용하여 미래의 감성을 예측하려는 시도가 증가하고 있다. 감성은 비언어적인 표현으로 나타나며 각 소비자가 가지고 있는 정확한 감성을 측정하기 어렵지만, 온라인에서 언어적인 표현으로 축적된 감성은 일정 기준을 부여하면 측정이 가능하다고 할 수 있다. 이러한 기준은 다양하지만 사전에 정의한 감성 사전은 하나의 효과적인 측정 기준이라고 할 수 있다. 각 감성에 해당하는 단어를 정리하고 검출 빈도와 같은 요소로 감성의 정도의 측정한다. 감성 사전을 구축하기 전에 감성을 연구 목적에 맞게 분류하는 것

은 감성 연구에서 핵심적인 과정이다. 감성과학에서 감성을 분류하는 방식은 크게 ‘차원적 감성’과 ‘개별 범주적 감성’으로 나눌 수 있으며 기존 연구는 주로 감성을 차원적 감성으로 나누어 감성 분석을 수행하고 있다. ‘차원적 감성’은 학술적 기반이 있는 분류 방식으로 긍정, 활성화 정도와 같은 차원으로 나누어져 있다. 심리학의 여러 연구에서 증명된 방식이고 차원 간 양의 정도와 상관관계를 쉽게 측정할 수 있다는 장점이 있어 기업 감성 연구에서도 주로 ‘차원적 감성’에 기반을 두고 있다. 그러나 ‘차원적 감성’은 국민이 실제로 가질 수 있는 보편적인 감성을 측정하기에는 효과적이지 않다고 볼 수 있다. 일반인이 느끼는 감성은 차원을 기반으로 하는 것이 아니라 직관에 기반 한다. 가령, 긍정적이고 능동적인 감성이라고 학술적으로 정의된 감성을 일반인은 이해하기 어려울 것이다. 심리 연구가 학술적인 연구를 위해 연구하는 감성은 일반인의 감성과 차이가 있으며, 일반인은 차원이 아니라 ‘슬픔’, ‘행복’, ‘분노’로 감성을 표현한다. 뿐만 아니라, 차원을 기준으로 감성을 분류하는 것은 심리학에 대한 학술적인 기초가 필요하기에 전문가가 아니면 어려움이 있다. 위와 같은 한계점이 있기에 ‘차원적 감성’을 이용한 연구는 의의가 있지만 고객이 느끼는 감성을 측정하려면 ‘개별 범주적 감성’을 기반으로 감성 분석을 할 필요가 있다고 할 수 있다. ‘개별 범주적 감성’은 일반인을 대상으로 실험하여 감성 단어를 분류하고 전문가가 마지막 정리 단계에만 개입하는 것으로 일정 주관성이 개입되기는 하지만 보편적인 감성을 연구하는데 설명력이 있다고 볼 수 있다. 감성의 범주는 ‘슬픔’, ‘행복’과 같은 것으로 정의되며 상호 독립을 가정으로 하기에, 상호 간 관계를 계량화하기는 어렵다고 할 수 있

다. 그러므로 개별 범주적인 감성과 차원적인 감성의 연구는 상호 보완적인 측면이 있다고 볼 수가 있다.

심리학·언어학과 같이 감성을 연구하는 분야에서는 감성을 개별 범주로 분류하는 연구를 진행해왔다. 일반적으로 적게는 4개에서 많게는 9개의 범주를 정하여 감성을 나누어왔으며 이 연구 결과를 인공지능이나 사물인터넷과 같은 융합 기술 분야에 활용하고 있다. 본 연구에서는 감성을 ‘슬픔’, ‘분노’, ‘기쁨’, ‘놀람’, ‘공포’, ‘혐오’, ‘지루함’, ‘흥미’, ‘통증’의 9개의 범주로 구분하여 감성사전을 구축한 기존 선행연구를 기반으로 우리나라 10대 기업에 대한 감성분석 연구를 수행하였다.

연구 대상 데이터로 뉴스를 사용하였다. 뉴스 데이터는 수집하기 용이하고 양이 많다는 장점이 있으며 문법을 준수하려는 경향이 있어 감성 사전에 있는 단어가 비교적 쉽게 검출된다. 뉴스는 독자의 감성을 여러 각도에서 충족시킨다. 각 언론사는 국민의 입장을 스스로의 특성에 맞게 보도하고자 하고 내용 안에는 인터뷰가 많기에 뉴스에는 기업에 대한 다양한 국민의 감성이 포함 되어 있다고 할 수 있다. 이런 이유에서 감성 사전을 이용한 다른 여러 선행 연구에서도 뉴스 데이터를 사용하였다.

우리나라 대기업은 우리나라 경제에서 큰 역할을 맡고 있으며 10대 기업은 약 반세기 동안 기업 활동을 영위해왔다. 각 기업에 대해 국민은 관심을 가지고 있으며 축적된 뉴스 데이터는 양이 방대하여 국민 감성을 측정하는 데이터로서 충분하다고 할 수 있다. 또한, 10대 기업의 이슈에 국민이 민감하게 반응하기에 이슈에 따른 국민 감성의 변화도 측정할 수 있다고 보았다. 본 연구는 뉴스 데이터 분석을 통해 우리나라 10대

기업에 대한 9가지 범주의 국민 감성을 비교하고 각 감성 빈도에 대한 비율과 순위를 확인했다. 더불어, 본 연구에서 제안한 감성 측정 방식이 기업별·시점별 감성 비교에 유의한지 확인하고자 가설을 세우고 검증하였다. 국민이 10대 기업에 대해 각각 다른 감성을 가지고 있는지, 특정 사건이 고객의 감성에 영향을 미칠 수 있는지 알아보았다.

본 논문의 구성은 다음과 같다. 2장에서는 감성의 정의와 감성 분류에 대해 참조한 선행 연구를 밝히고 감성 분석을 기업 연구에 활용한 연구를 살펴보았다. 3장에서는 개별 범주적 감성을 이용한 감성 사전이 기업별·시점별 감성 분석에 유의한지 확인하는 가설을 제시하고 가설을 제시한 배경에 대해서 설명하였다. 4장에서는 연구 데이터의 특성과 연구 절차를 설명하며 5장에서는 연구 결과를 검증하고 시사점을 제시하고자 한다. 마지막 6장에서는 결론과 연구의 한계점을 설명하고 보완점과 향후 연구 방안에 대해서 제시한다.

2. 선행 연구

2.1 기업 분석에 활용하는 감성의 정의

‘감성’은 사전적인 정의로서 인간이 오관을 통해 상황이나 사물에서부터 받은 자극 혹은 인상인 ‘감각’을 지각하여 표상을 형성하는 인간의 인식 능력이다(Song, 2016). ‘감정’ 혹은 ‘정서’는 물리적 자극뿐만 아니라 어떤 현상이나 일에 대하여 일어나는 마음이나 기분을 포함하는 의미로 단어가 서로 구분된다(Song, 2016). 영어에서 ‘감성’을 의미하는 단어는 ‘sensitivity’라고 할 수

있으며 ‘감정’을 의미하는 단어는 ‘emotion’이나 ‘feeling’이라고 할 수 있다. 한글에서보다 영어에서 ‘감성’과 ‘감정’을 사전적으로 구분하기가 어려우며 혼용해서 사용하는 경우가 많다. 우리나라 국책연구에서 감성공학을 영어로 표기할 때는 ‘Sensibility Ergonomics’라는 단어를 사용한다 (Song, 2016).

‘감성’과 ‘감정’을 구분하는 것이나 ‘정서’, ‘감각’과 같은 유사 단어 간의 의미를 정확히 파악하는 연구는 존재하지만 기업 분석이라는 목표에 적합하지 않다. 1997년 설립된 ‘한국감성과학회’는 ‘감성’의 범주에 ‘감정’, ‘정서’와 같은 기타 유사 단어의 범주를 모두 포함하고 있다. 감성 어휘를 추출하는 한 선행 연구(Song, 2016)는 감성 단어에 감정을 비롯한 유사한 범주에 해당하는 어휘를 모두 포함하는 경향이 있다고 밝혔다. 기업 분석에는 유사 단어의 의미를 구분하지 않고 모두 사용하는 것이 적절하다고 할 수 있다.

2.2 한국어 감성 단어 범주화

감성 단어를 추출하고 범주화하고자 하는 시도는 심리학·언어학 분야에서 주로 이루어지다가 최근에는 융합 공학을 비롯한 다방면에서 연구가 이루어지고 있다. 감성은 사건에 대한 개인의 평가이며, 이러한 감성은 표정과 같은 비언어적인 표현과 말과 같은 언어적인 표현으로 나타나며 생리 신호나 신체의 전기 신호로 나타나기도 한다(Sohn et al., 2012). 온라인에서 측정할 수 있는 빅 데이터는 여러 형태가 있으나 영상에 담겨 있는 비언어적인 표현을 측정하기에는 아직 한계가 있다. 사람이 감성을 표현하는데 있어 비언어적인 표현에 더 비중을 둔다는 한계점이 있

지만 비언어적인 표현은 계량화하기 어렵다는 단점에 비해, 언어적인 표현으로 표현한 감성은 객관화하기 용이하다. 언어적 표현은 가장 정교하고 단호한 표현 수단으로서 감성의 구조를 이해하고 단어를 범주화하는 것은 의의가 있다고 할 수 있다(Sohn et al., 2012). 이러한 연구를 기업 분석에 응용하는 시도는 기업이 국민 감성을 이해하고 예측하는데 의의가 있다고 할 수 있다.

국내에서는 1990년대부터 한국어 감성 단어를 분류하는 연구가 시작되었다. 영어권 연구를 기초로 하여 1,340개의 단어 중 213개의 감성 단어가 선별되었다(Ahn et al, 1993). 감성단어는 성격, 행동, 감성 상태, 신체 감각을 기반으로 추출하였으며 감성 구조는 쾌-불쾌의 단일 차원을 갖는 것으로 밝혔다. 2000년대에 들어 한국어 단어의 감성 구조를 파악하는 연구는 안신호 등(1993)이 구축한 감성 단어 목록을 정교하게 보완하고자 했다(Park et al, 2005). 이들은 연세대학교 언어연구소에서 출판한 ‘연세대학교 어휘총집’을 기초로 사용 빈도가 높은 단어를 주로 선정하였다. 64,666개의 단어를 신체 감각, 성격, 감성 상태, 행동, 동기, 상태, 관계, 구, 기타 8가지 방법을 사용하여 총 434개의 단어 목록을 완성하였다. 이들은 감성 구조를 쾌-불쾌, 타인 초점적-자기 초점적 차원이라고 밝혔다.

차원적 감성 구조에 한계를 느낀 이후 연구에서는 박인조 등(2005)의 연구에 기반을 두고 감성 단어 목록을 개별 범주로 나누었다(Sohn et al, 2012). 이들은 차원에 의한 감성 구조는 직관적이지 않고 분류에 혼란이 필요하기에 보편적인 감성을 측정하는데 한계가 있다고 밝혔다. 차원을 기반으로 한 감성 구조는 감성을 평가, 활성화, 통제, 힘 등의 차원으로 나눈다. 예를 들어, ‘평가’는 인간이 감성을 얼마나 긍정/부정으로

느끼는지 평가하며 ‘활성화’는 인간이 특정 감성 상태에서 능동적/수동적인 행동을 취하는 경향을 평가한다. 차원으로 감성 구조를 파악하는 것은 학술적으로 증명된 방식이며 각 차원 간의 관계를 계량화하기 용이하다는 장점이 있다. 그러나 일반인은 감성을 차원으로 나누어 느끼지 않기에 복잡한 감성을 차원으로 나누는 것은 오차가 크다고 할 수 있다. 이러한 측면에서 감성 구조를 개별 범주적 감성으로 정의하는 것은 차원적 감성의 한계점을 극복하는 가장 오래된 방식이라고 할 수 있다(Greenwald et al., 1989). 차원적 관점은 감성을 연속적인 차원 상에 위치하는 값으로 정의하는데 반해, 개별 범주적인 감성은 각 감성을 독립적인 감성의 범주로 정의한다. 개별 범주적 감성의 가장 대표적인 예는 에크만(Ekman)이 정의한 기본 감성이며 ‘행복’, ‘슬픔’, ‘공포’, ‘분노’, ‘혐오’, ‘놀람’이다. 이러한 분류는 일상생활에서 사람들이 감성을 묘사하는 형태와 유사하고 직관적이어서 이해하기가 쉽다(Sohn et al., 2012). 6개의 감성에 HCI (Human Computer Interface) 분야에 주로 활용되는 감성 중 활용도가 높은 ‘흥미’, ‘지루함’, ‘통증’을 감성 범주에 포함하여 총 9가지로 감성을 범주화하였다. 이 연구는 선행 연구에서 감성 사전의 데이터로 주로 활용한 ‘연세대학교 어휘 총집’을 기초로 기타와 중립의 감성을 포함한 11개의 감성 범주에서 최종 504개의 단어를 추출하였다. 충남대학교 뇌과학 연구소의 6명의 연구원이 1차에 걸쳐 64,666개의 단어 중에서 1,228개의 감성 단어를 추출하였고, 2차에는 80명의 청주대학교 대학생이 연구에 참가하여 각 단어의 감성과 강도에 대한 설문을 통해 최종 504개의 감성 단어를 추출하였다. 각 단어를 가장 높은 감성을 나타내는 범주로 나누었다. 두 번째 감성 강도가 첫 번째

감성 강도의 80%비율에 달했을 시에는 단어 당 감성의 범주를 2가지로 부여하였으며, 마찬가지로 세 번째로 높은 감성에도 같은 방식을 취하였다. 그러나 네 번째부터는 같은 방식을 취하지 않았다.

감성 단어 목록에는 한국어 표준어 규정과 맞춤법을 따르지 않는 단어도 있다. 이는 연구의 기반이 된 연세대학교 한국어 어휘 빈도 자료집이 최근 10년간 발행된 출판물과 영상물에 기초하고 있기 때문이다. 이 기준을 적용한 결과 504개의 감성 표현단어 중 한 가지 감성으로만 구분되는 426개의 단어는 ‘슬픔’, ‘분노’, ‘기쁨’, ‘놀람’, ‘공포’, ‘혐오’, ‘지루함’, ‘흥미’, ‘통증’ 순으로 많았다.

이 연구에는 몇 가지 한계점이 있다. 7명의 전문 연구원과 일반적인 감성을 분류하는 80명의 대학생 표본이 크지 않기에 보편적인 감성을 일반화할 수 있다는 한계가 있다. 기존에 차원적 감성 구조를 기반으로 수행한 선행연구에서는 감성을 신체감각, 성격, 감성 상태 등의 학술적으로 분류가 가능한 여러 기준으로 적용했다면 이 연구는 일상생활에서의 활용 여부와 감성 표현과의 관련성 여부에 집중하여 범주화하였기에 일정 주관성이 개입되었을 수 있다. 감성 사전의 기반 데이터인 ‘연세대학교 어휘 빈도 사전’의 실용성과 정확도도 보장할 수는 없기에 본질적인 한계점이라고 할 수 있다.

2.3 빅 데이터와 감성 분석을 활용한 기업 연구

감성 분석에는 여러 기법이 있지만 빅 데이터와 감성 분석을 기업 연구에 활용할 때 사용하는 대표적인 기법은 기록된 자연어에 포함된 감성

을 평가하고 분류하는 것이다(Jung et al., 2015). 선행 연구는 주로 빅 데이터에 존재하는 감성 단어를 ‘평가’, ‘활성화’ 방식과 같은 차원적 감성 구조로 분류하고 와 기업 관련 뉴스 데이터와 해당 기업의 초과이익률의 관계를 분석하여 기업 성과를 예측하고자 하였다. 감성 사전을 구축하는 것은 각 차원을 측정하는 중요한 기준으로 분석의 효과에 중요하다고 할 수 있다. 감성의 정도를 측정하는 기준은 목적에 따라 다양하지만 감성사전에 수록된 단어의 출현 빈도로 감성을 파악하는 것이 가장 보편적이라고 할 수 있다.

감성 사전을 구축하는 방법은 전문 연구원이거나 설문조사를 통하여 사전에 정의하는 방법이 있다. 군집화를 통하여 유사성이 높은 단어나 거리가 가까운 단어를 기반으로 유사 단어의 집단을 특정 감성으로 구축하기도 한다. 감성 단어는 형용사, 명사 단위로 끊기도 하며 형태소 단위인 경우도 있다. 기업 연구에서는 언어학적·심리학적 방법으로 규칙을 만들거나 특정 기업에 맞게 식을 만들어 감성을 분류하고 감성 점수를 산출하기도 한다. 시기나 조직에 따라서 같은 단어라도 다른 감성을 적용하기도 하며 연구의 성능을 위해 기업별로 개별적인 감성사전을 조직화하기도 한다. 통일적인 감성 사전의 구축과 개별적인 감성 사전의 구축은 기준의 표준화와 검색 효과 사이에서 장단점이 있다고 할 수 있다.

차원적 감성 구조에서 기업 분석에 주로 사용하는 차원은 ‘평가’차원으로 고객의 긍/부정 감성을 측정한다. 판매량이나 가격과 같은 기업 요소를 감성의 긍/부정의 강도·상관관계와 결합하여 예측하는 연구도 다소 있다. 이러한 연구는 감성과 경영 성과의 관계를 계량화했다는 점에서 의의가 있지만 감성을 긍/부정으로만 나눌 수가 없기에 한계점이 존재한다고 할 수 있다.

본 연구에서는 선행연구인 ‘한국어 감성 단어 범주화와 감정 분석 연구’를 바탕으로 우리나라 10대 기업에 대한 국민 감성을 분석해보려고 한다. 보편적으로 나눈 9가지 감성을 국민들은 어떻게 갖고 있는지 파악하고, 기업별로 국민의 감성은 차이가 있는지, 주요 이슈에 따라 시점별로 감성은 어떻게 변화하는지 측정하는 연구를 수행하려고 한다. 일반 국민이 느낄 수 있는 기업에 대한 보편적인 감성을 보여주는 데 중점을 두었다.

3. 연구 가설

산업 혁명 이후 20세기 중후반까지도 기업에게 상품을 결정하는 요소는 품질과 가격이었다고 할 수 있다. 마이클 포터(Michael Porter)는 시장 본원적 전략을 제시하며 기업의 성공은 기본적으로 가격과 품질에 달려있다고 주장했다. 그러나 가격과 품질이 표준화되고 경쟁이 심화되면서 기업은 다른 요소도 고려하기 시작했다. 기업은 상품 자체를 중시하는 기업 중심의 사고에서 벗어나 소비자 입장에서 상품을 생각하는 고객 중심의 사고를 할 필요를 느꼈다. 기업은 감성, 이미지, 경험과 같은 비기능적인 가치를 추구하는 새로운 전략을 모색하였고 감성은 기업이 고객을 이해하고 더 높은 성과를 내기 위한 핵심 요소가 되었다(Song, 2016). 기업은 감성공학을 적극적으로 활용하여 기업의 성과를 극대화할 방법을 모색하였다. 통신 매체와 소셜 미디어의 발달로 고객과 기업은 시공간을 초월하여 소통을 할 수 있게 되면서 이제 고객은 단순히 소비자가 아니라 기업을 홍보하는 또 다른 내부고객으로서의 역할을 동시에 하고 있다. 뿐

만 아니라, 상품을 구매하지 않은 고객은 온라인에서 집단화된 감성에 영향을 받으며 자신의 감성을 다른 고객에게 공유한다. SNS나 양방향성 소통 매체인 인터넷 통신매체의 발달 전에는 기업이 소비자의 감성을 측정하기 위해 단편적인 설문조사나 추측에 의해서 일방적으로 소통했지만 이제는 시간과 공간을 초월한 소통이 가능해지면서 기업은 고객의 감성을 즉시 이해해야 할 필요성을 느꼈다(Kotler, 2010). 소비자는 각 기업에 대한 집단적인 감성을 가질 수 있으며 이러한 감성은 집단적인 국민 감성을 구축한다. 국민 감성을 사회적 감성(social sentiment)이라고 표현하는 다른 연구도 있다. 집단화된 국민 감성은 사건에 따라 빠르게 변화하며 그 효과는 강력하다고 할 수 있다. 2016년 옥시의 가슴기 살균제 사건이나 2015년 폭스바겐의 자동차 연비·가스 점검 조작 사건은 국민의 감성이 온라인상에서 집단화된 대표적인 사례이다. 온라인에서 집단화된 감성은 기업의 실적에 즉시 영향을 미치며 미래의 실적에도 영향을 미칠 수 있다. 단순히 상품을 구매한 소비자로서의 반응이 아니라 사회 집단적인 반응으로 이 효과는 매우 크기 때문에 기업은 국민 감성을 인지하고 계량적인 데이터로 나타낼 필요성이 있다고 할 수 있다. 결과적으로, 기업은 타 회사와의 감성 비교를 통하여 혹은 시점별 감성 비교를 통하여 감성을 측정하고 예측하여 전략적인 선택을 할 필요성이 있다.

우리나라 10대 기업은 세계에서 유례가 없는 50년 이상 시장을 독보적으로 점유한 직계 가족 승계형 기업이며 ‘Jaebul’이라는 신조어로 영어 사전에 등재 되어있다. 10대 기업에는 확고한 경영철학이나 경영자의 가족 문화가 스며들어 있어 국민은 각 기업마다 구별되는 감성을 가지고

있다고 할 수 있다. 우리나라의 경제는 10대 기업에 크게 의존하고 있으며 대중의 10대 기업에 대한 관심도가 높다. 사건 사고마다 기업의 시가 총액과 같은 기업 가치가 크게 변할 수 있다. 예를 들어, ‘포스코’는 라면 사건으로 큰 폭의 주가 하락을 경험했다. 10대 기업에 대한 국민의 관심도가 높아 관련된 다량의 뉴스 데이터는 국민 감성을 적절히 효과적으로 측정할 대상이라고 보았다. 이상의 논의를 정리하며 다음의 2가지 가설을 제시한다.

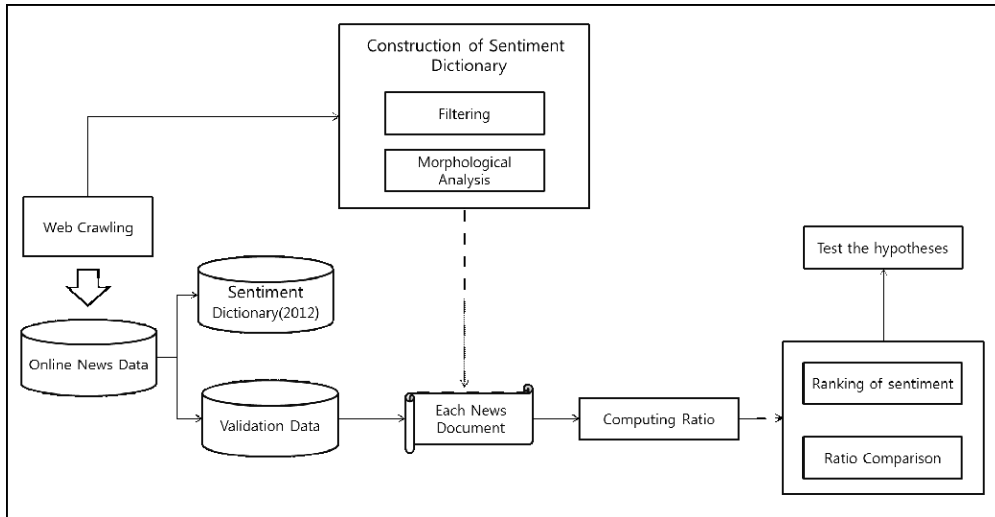
H1: 기업에 따라 국민의 감성이 다르다.

H2: 특정 사건은 국민의 감성에 영향을 미친다.

4. 연구 방안

4.1 연구 절차

본 연구는 다음과 같은 연구 절차에 따라 수행되었다(<Figure 1> 참조). 온라인 포털 서비스를 제공하는 ‘네이버’의 증권 뉴스 게시판에서 기업 뉴스를 웹 크롤링(web crawling)하였다. 2014년 1월부터 2016년 1월까지의 25개월 간 축적된 기업관련 뉴스를 수집하여 감성 분석 대상 데이터로 선정하였다. 감성 사전의 단어 목록은 선행연구인 ‘한국어 감성표현단어의 추출과 범주화’ 연구에서 추출한 504개의 단어를 기반으로 구축하였다(Sohn et al., 2012). 대상 데이터와 감성 사전 데이터를 각각 R에서 패키지로 제공하는 ‘tm’과 ‘KoNLP’를 활용하여 분석을 진행하였다. ‘KoNLP’의 명사 추출 기능을 활용하여 명사 위주로 분석을 수행하였으며 특수문자·영어·숫자



〈Figure 1〉 Research process using big-data and multi-categorical sentiment dictionary

및 기타 방해요소는 제외하였다.

먼저, 각 기업의 뉴스 데이터에서 감성 단어가 검출된 빈도를 확인하고 비율로 바꾸어 감성의 분포를 확인하였다. 감성의 비율이 높은 순서대로 순위를 부여하여 비교하였으며 10대 기업 중 데이터의 수가 더 많고 경제에 더 큰 영향을 끼치는 4대 기업은 따로 분리하여 비교하였다.

첫 번째 가설 검정으로, 기업 간 국민의 감성이 유의미하게 차이가 있는지 확인하기 위해 2개 기업을 쌍으로 검정하는 대응표본 T검정을 실시하였으며 순위적으로 유의한지 확인하기 위해 Duncan 검정을 실시했다.

두 번째 가설 검정으로, 국민의 감성이 특정 사건의 발생으로 변화가 발생하는지 검증하기 위해 대한항공 사례를 활용했다. 대한항공의 ‘땅콩회항’ 사건이 일어난 2014년 12월 감성 단어 검출 수가 월 평균 감성 단어 검출 수와 비교하여 얼마나 증가했는지 알아보았다. 감성을 긍/부정 감성으로 다시 나누어 긍/부정 감성의 변화

추이를 사건 전후로 확인하고 시점에 따라 해석하였다. 사건 전후 각 범주의 감성 분포를 비교하여 사건이 국민의 감성에 어떤 영향을 미쳤는지 해석하였다. 추가적으로 사건 전후에 급증한 감성 중 ‘분노’의 감성을 선택하여 검출된 단어의 빈도와 단어 목록의 변화를 구체적으로 알아보았다. 이러한 단어 목록의 변화는 국민의 감성 변화를 구체적으로 증명할 설명력이 있는 요소라고 보았다.

4.2 연구 데이터

본 연구는 국내 점유율 1위 포털 사이트인 ‘네이버’가 제공하는 증권 서비스 중, 기업별 뉴스 게시판에서 2014년 1월부터 2016년 1월까지 데이터를 월 별로 나누어 25개월의 데이터를 활용하였다. 총 90,683개의 데이터를 10대 기업에 대한 국민 감성을 보여주는 검증 데이터로 사용했다. 분석 대상 기업의 선정은 우리나라 공정거래

위원회에서 2015년 지정한 상호출자·채무보증 제한 기업 집단 중 자산총액 순위를 기준으로 상위 10개 기업을 선정하였다. 국가 산업을 운영하고 있는 공기업이나 공익 산업의 성격을 지니고 있는 금융회사는 제외하였다. 공기업으로는 한국전력·한국토지주택공사·한국도로공사·한국가스공사가 제외되었고, 금융기업으로는 농협이 위와 같은 이유에서 제외되었다. 공정거래위원회는 2015년 우리나라 10대 기업을 선정하는데 있어서 동일한 기준을 사용하였으며 주요 정보를 공시할 때마다 위의 기준을 활용하고 있다. 기업 그룹의 모든 데이터를 수집하여 분석하는데에는 한계가 있으므로 상위 10대 기업 그룹 중 그룹을 대표하는 계열사를 선정하였다. 대표 기업으로 가장 화제가 되는 기업을 선정하였는데, 그 이유는 시가 총액이나 매출액은 미래 상황을 반영하고 변동성이 크기 때문에 뉴스 데이터의 총량을 기준으로 하여 대표 기업을 선정하는 것이 더 적절하다고 보았기 때문이다. 이에 따라 자산총액 기준에서 선정된 기업 집단은 순서대로 ‘삼성’, ‘현대자동차’, ‘SK’, ‘LG’, ‘롯데’,

‘GS’, ‘현대중공업’, ‘한진’, ‘한화’이다. 위의 기업 집단에서 화제성으로 선정된 각 그룹의 대표 기업은 ‘삼성전자’, ‘현대자동차’, ‘포스코’, ‘SK 텔레콤’, ‘LG전자’, ‘롯데쇼핑’, ‘GS건설’, ‘현대중공업’, ‘대한항공’, ‘한화’이다(<Table 1> 참조).

4.3 감성 사전 구축

감성사전은 선행연구인 ‘한국어 감성표현단어의 추출과 범주화’ 연구에서 추출한 504개의 단어를 사용하였다(Sohn et al, 2012). ‘KoNLP’는 명사 추출에 특화되어 있기에 명사 위주로 추출하였으며 명사 형태가 아닌 것은 명사 형태로 만들었다. 꼭 명사가 아니더라도 ‘KoNLP’를 통해 나온 결과가 의미가 성립이 되는 단어라면 검색될 수 있는 형태로 만들었으며 대부분의 단어가 어간만을 포함하고 있다. 단어는 일부 예외를 제외하고는 2글자 이상의 단어이며 중의적 의미를 가진 단어는 제외하였다. 예를 들어서 놀라움 영역의 ‘이상’이라는 단어는 ‘이상적이다’, ‘이상하다’로 2가지 의미를 가지고 있어 정확하게 구분

<Table 1> Korean top-10 conglomerates sorted by asset amount (Fair Trade Commission, 2015)

	Asset (Trillion won)	Affiliates (unit)	Representative Affiliate	The amount of articles (unit)
Samsung	351.5	67	Samsung Electronics	31,681
Hyundai Motors	194.1	51	Hyundai Motors	17,252
SK	152.4	82	SK Telecom	9,627
LG	105.5	63	LG Electronics	9,384
Lotte	93.4	80	Lotte Shopping	2,824
Posco	84.5	51	Posco	7,580
GS	58.5	79	GS Engineering & Construction	2,144
Hyundai Heavy industries	57.5	27	Hyundai Heavy industries	4,141
Hanjin	38.4	46	Korean Air	3,874
Hanhwa	38	52	Hanhwa	2,176

할 수 없었기에 제외하였다. 감성 사전 단어 목록에서 띄어쓰기 및 기타 분석을 방해하는 요소는 제거하여 오차를 최소화하고자 하였다. 두 가지나 세 가지 감성의 범주에 중복하여 속해 있는 단어는 각 범주에 모두 포함시켰다. 선행 연구에서 제시한 감성 단어 목록을 검색이 용이하도록 여러 형태로 만들었기 때문에 일부 단어가 제외되었음에도 오차려 증가하였다. 위의 과정을 통하여 구축한 감성 사전의 단어 개수는 총 580개이다. 각 감정 범주 당 단어 수는 ‘슬픔’ 157개, ‘분노’ 120개, ‘행복’ 101개, ‘혐오’ 53개, ‘놀라움’ 47개, ‘공포’ 44개, ‘흥미’ 25개, ‘지루함’ 21개, ‘고통’ 12개이었다(<Table 2> 참조). 감성사전은 개별 기업마다 구축하지 않고 통일적으로 적용하였다.

<Table 2> Identification of 580 feeling words of single sentiment

Sentiment	Number of Words
Sadness	157
Anger	120
Happiness	101
Disgust	53
Surprise	47
Fear	44
Interest	25
Boredom	21
Pain	12

4.4 감성 비율 분포와 감성 순위

각 기업의 감성 순위를 부여하고 기업 간 감성의 차이가 있는지 비교하기 위하여 감성을 비율로 나타내었다. 전체 감성의 분포를 확인하고 순위를 정하는데 사용된 기업의 전체 감성 비율은 ‘해당 감성 단어 검출 빈도 / 전체 감성 단어 검

출 빈도’로 측정하였다(<Formula 1> 참조).

$$Sentiment\ ratio = \frac{Terms\ of\ single\ sentiment}{Total\ terms\ of\ sentiment} \quad (1)$$

국민의 감성이 기업별로 차이가 있는지 검정하기 위해 감성 비율을 월별로 구분하여 총 25개의 표본으로 만들었다. 가설 검정에 사용된 월별 감성 비율은 ‘월별 특정 감성 단어 빈도/월별 전체 감성 단어 빈도’이다(<Formula 2> 참조).

$$Monthly\ sentiment\ ratio = \frac{Monthly\ terms\ of\ single\ sentiment}{Monthly\ total\ terms\ of\ sentiment} \quad (2)$$

기업별 감성 순위는 해당 감성의 비율이 더 높은 순으로 내림차순으로 정렬하였다.

10대 기업 중에서도 우리나라 경제에 더 큰 영향을 미치고 있고 국민의 관심이 더 집중되어 있다고 할 수 있는 4대 기업은 따로 비교하였다. 4대 기업은 ‘삼성’, ‘현대자동차’, ‘SK’, ‘LG’로 이들 간의 차이는 별도로 가시화하여 비교하였다.

4.5 가설 검증 방안

4.5.1 가설 1: 기업 간 감성 차이 증명의 연구 절차

‘기업에 따라 국민의 감성이 다르다’라는 가설을 검정하기 위해 두 가지 통계적인 검증 방법을 사용하였다.

첫 번째 통계적 검정으로 두 기업의 감성을 쌍으로 비교하는 대응표본 T검정을 수행하였다. 본 논문에서 10개 기업 9개 범주 감성의 대응표본 T검정 결과를 모두 제시하기에는 한계점이 존재하므로 대표적으로 기업 두 쌍을 선택하여 검정을 실시하였다. 1위 기업으로 가장 긍정적인

감성을 가질 것이라고 예상되는 ‘삼성’과 최근까지도 부정적인 사건을 겪어 부정적인 감성을 가질 것이라고 예상되는 ‘한진’을 첫 번째 쌍으로 선택하여 비교하였다. 비슷하게 부정적인 사건을 겪은 ‘SK’와 ‘한진’을 두 번째 쌍으로 검정을 수행하여 감성에 차이가 있는지 살펴보았다.

두 번째 통계적 검정으로 감성이 순위적으로 유의한 차이가 있는지 증명하는 Duncan 검정을 수행하였다. 대표적으로 감성 사전에서 단어 수가 가장 많은 ‘슬픔’을 선택하여 10대 기업의 ‘슬픔’ 감성이 순위적으로 유의하게 다른지 검정하고 해석하였다.

4.5.2 가설 2: 시점 간 감성 차이 증명의 연구 절차

두 번째 가설인 ‘특정 사건이 국민의 감성에 영향을 미친다.’를 증명하기 위해 ‘대한항공’의 ‘땅콩회항’사건이 일어난 2014년 12월과 나머지 24개월 데이터에서 검출된 감성 단어를 비교 분석하였다. 전체 감성 단어의 평균 검출 빈도에 대하여 12월 감성 단어의 관측치가 어떤 특이점이 있는지 밝혔다.

추가적으로 감성의 개별 범주를 긍/부정으로

다시 나누어 시점별 변화 추이를 확인하고 시사점을 해석하였다. 심리학자인 에크만(Ekman)이 정의한 6가지 기본 감성 범주 중 4가지 범주인 ‘슬픔’, ‘분노’, ‘혐오’, ‘공포’를 부정으로 분류하였고 ‘행복’을 긍정으로 분류하여 검출된 단어의 빈도 변화를 확인했다. 감성을 개별 범주로 나누고 긍/부정으로 다시 나누는 시도는 선행 연구에서 타당성이 입증되었다. 위와 같은 분류는 심리학계의 선행 연구와 비슷한 방식이라고 볼 수 있다(Lee et al, 2008). 또한, ‘땅콩회항’이 일어난 달과 전 달의 전 범주 감성 단어 빈도를 비교하여 국민의 감성 변화를 총체적으로 알아보았다. 증가한 감성 중에 ‘분노’를 선택하여 검출된 단어 목록의 변화를 알아보고 국민이 ‘분노’의 감성을 어떠한 단어로 표현하였는지 구체적으로 알아보았다.

5. 연구 결과

5.1 각 기업의 감성 비율과 감성 순위

기업별 감성 비율을 <Formula 1>, <Formula

<Table 3> Sentiment ratio

	Samsung	Hyundai Motors	SK	LG	Lotte	Posco	GS	Hyundai Heavy industries	Hanjin	Hanhwa
Sadness	25.70%	23.86%	25.96%	20.68%	30.16%	29.64%	32.21%	34.35%	26.82%	30.66%
Anger	9.83%	9.67%	8.60%	7.58%	6.81%	9.52%	10.85%	9.96%	10.18%	5.81%
Happiness	18.61%	19.50%	13.08%	21.52%	20.38%	12.07%	17.35%	15.17%	24.15%	24.24%
Disgust	17.50%	20.94%	20.23%	15.18%	19.55%	25.93%	23.64%	21.30%	16.81%	17.28%
Surprise	7.81%	7.23%	4.94%	7.12%	2.71%	7.92%	5.31%	6.88%	3.97%	4.97%
Fear	7.05%	8.92%	5.18%	3.94%	7.11%	9.15%	6.83%	7.10%	5.20%	4.97%
Interest	6.00%	4.43%	3.68%	14.86%	6.22%	3.78%	1.41%	2.51%	2.70%	3.98%
Boredom	5.33%	3.69%	14.20%	6.98%	5.03%	1.13%	1.74%	1.01%	5.72%	6.50%
Pain	2.16%	1.78%	4.13%	2.13%	2.02%	0.86%	0.65%	1.72%	4.45%	1.61%

〈Table 4〉 Sentiment ranking

Rank	Sadness	Anger	Happiness	Disgust	Surprise	Fear	Interest	Boredom	Pain
1	Hyundai Heavy Industrials	GS	Hanhwa	Posco	Posco	Posco	LG	SK	Hanjin
2	GS	Hanjin	Hanjin	GS	Samsung	Hyundai Motors	Lotte	LG	SK
3	Hanhwa	Hyundai Heavy Industrials	LG	Hyundai Heavy Industrials	Hyundai Motors	Lotte	Samsung	Hanhwa	Samsung
4	Lotte	Samsung	Lotte	Hyundai Motors	LG	Hyundai Heavy Industrials	Hyundai Motors	Hanjin	LG
5	Posco	Hyundai Motors	Hyundai Motors	SK	Hyundai Heavy Industrials	Samsung	Hanhwa	Samsung	Lotte
6	Hanjin	Posco	Samsung	Lotte	GS	GS	Posco	Lotte	Hyundai Motors
7	SK	SK	GS	Samsung	Hanhwa	Hanjin	SK	Hyundai Motors	Hyundai Heavy Industrials
8	Samsung	LG	Hyundai Heavy Industrials	Hanhwa	SK	SK	Hanjin	GS	Hanhwa
9	Hyundai Motors	Lotte	SK	Hanjin	Hanjin	Hanhwa	Hyundai Heavy Industrials	Posco	Posco
10	LG	Hanhwa	Posco	LG	Lotte	LG	GS	Hyundai Heavy Industrials	GS

2>의 방식으로 산출하고 각 감성의 비율이 높은 순서대로 내림차순으로 정렬하였다(〈Table 3〉, 〈Table 4〉 참조). 각 감성에 따른 빈도는 감정 사전에 등록되어 있는 단어 수에 영향을 받을 수 있기에 절대적인 지표로 활용되기보다는 상대적인 지표로서 기업별·시점별 비교에 활용하는 것이 바람직하다고 할 수 있다.

각 감성별 1위 기업은 다음과 같다. 슬픔은 ‘현대중공업’, 분노는 ‘GS’, 행복은 ‘한화’, 혐오는 ‘포스코’, 놀라움은 ‘포스코’, 공포는 ‘포스코’, 흥미는 ‘LG’, 지루함은 ‘SK’, 고통은 ‘한진’이 1위를 차지했다.

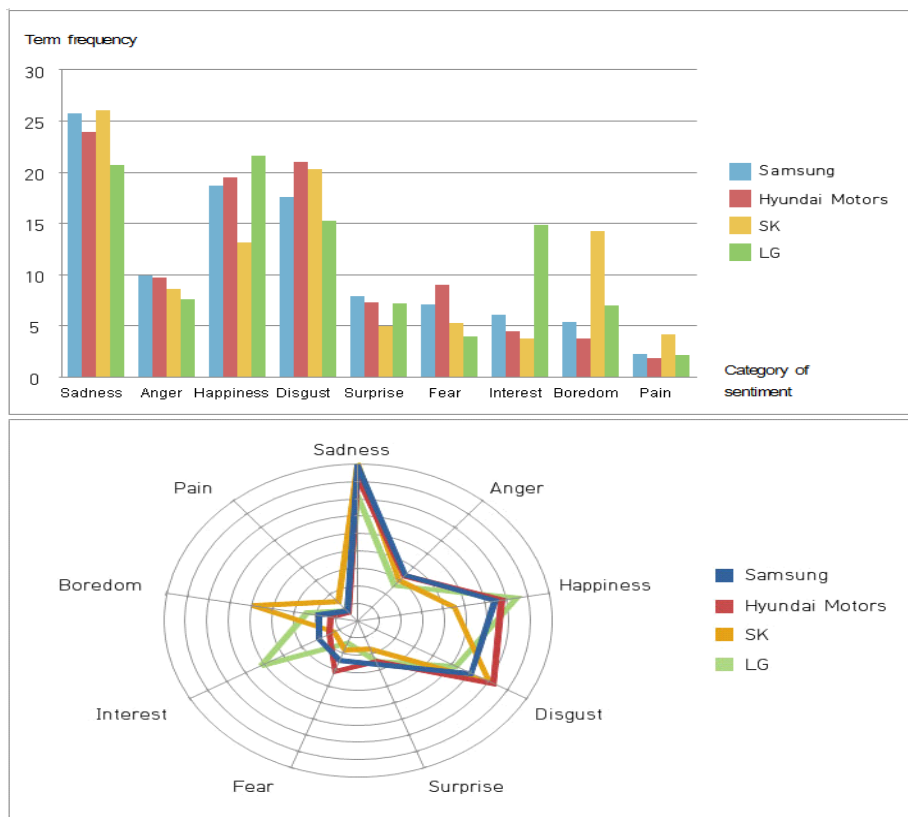
감성 순위를 통해 제시할 수 있는 시사점은

다음과 같다. 첫째로, 기업에 대한 감성은 주력 산업의 시장 상황에 의해 크게 영향을 받을 수 있다. 최근 선박 시장의 부진으로 대량으로 정리해고를 실시한 ‘현대 중공업’이 ‘슬픔’ 영역에서 1위를 차지했다. 중국 철강 기업의 강세로 상대적으로 부진하고 있는 ‘포스코’가 ‘혐오’, ‘놀라움’, ‘공포’ 영역에서 1위를 차지했다. 우리나라 건설 시장의 포화로 수익성이 떨어진 ‘GS건설’도 ‘분노’ 영역에서 1위를 차지했다. 둘째로, 긍정적인 감성과 부정적인 감성은 동시에 존재할 수가 있다. 개별 범주적 감성은 각 감성이 독립이라는 가정이 있기 때문에 ‘행복’의 순위가 높다고 해서 반대인 것처럼 보이는 ‘분노’

나 ‘슬픔’이 반드시 낮아야 하는 것은 아니다. 이러한 점은 예를 통하여 알 수 있는데 ‘한진’은 ‘분노’와 ‘고통’의 순위가 각 2위, 1위이지만 ‘행복’의 순위는 2위였다.

또한, 부정적인 감성인 ‘혐오’는 9위로 각 감성 간의 상관관계가 확연하게 있다고 보기에 어려웠다. 그러나 분노 순위가 가장 낮은 ‘한화’가 ‘행복’ 순위가 가장 높고, ‘협오’와 ‘공포’의 감성이 가장 높은 ‘포스코’가 ‘행복’ 순위가 가장 낮은 것을 보면 감성 간 관계가 일정 부분 있다고 볼 수도 있다. 셋째로, 비슷한 성격의 사건이라도 국민의 감성은 국민의 삶에 직접적으로 영향을 줄 수 있는 사건에 더 민감하게 반응한다.

‘대한항공’의 ‘땅콩회항’ 사건은 국민의 삶에 바로 영향을 줄 수 있는 사건이라고 할 수 있다. 월별 감성 비율과 전체 감성 순위의 움직임을 기반으로 연구한 결과, 일련의 부정적인 사건인 ‘롯데’의 경영 승계 문제나 ‘SK’ 최태원 회장의 외도 문제는 기업의 감성에 크게 영향을 미치지 않았다고 판단했다. 감성 단어의 검출 빈도가 사건이 발생한 달에 크게 증가하지 않았고 긍/부정 감성의 움직임에도 큰 변화가 없었다. 그 이유는 소비자가 재화나 서비스를 이용하는데 있어 ‘롯데’나 ‘SK’ 일가의 사적인 사건에 부정적인 영향을 받지 않았기 때문이라고 추측된다. 하지만 ‘대한항공’의 경우에는 회항 사건으로 인해 일부



〈Figure 2〉 Sentiment words term frequency of top 4 conglomerates

소비자가 직접 이용에 불편을 겪었다. 불편 사항을 겪었음에도 회사는 보상이나 사과를 하지 않았기에 ‘대한항공’을 이용하는 다른 국민도 잠재적 피해자가 될 수 있는 상황이었다. 사무장이나 승무원과 같은 피고용인도 이 사건에 의해 피해를 입은 반면에 ‘롯데’나 ‘SK’는 피고용인이나 소비자에게 피해를 입힌 바가 없기에 국민의 입장에서 느끼는 의미는 다를 수 있다고 보았다.

4대 기업을 분리해서 비교 분석을 한 결과, 전 범주에서 ‘삼성’과 ‘현대자동차’는 감성 비율의 차이가 크지 않았다(<Figure 2> 참조). 반면에 ‘SK’는 긍정적인 감성이라고 할 수 있는 ‘행복’의 비율이 낮았고 부정적인 감성인 ‘지루함’의 비율이 상당히 높고 ‘고통’의 비율이 다소 높았다. ‘SK’의 ‘슬픔’, ‘혐오’의 비율도 비교적 높은 것으로 나타났으며 ‘놀람’의 감성은 비교적 낮은 것으로 나타났다. ‘LG’는 긍정적인 감성이라고 할 수 있는 ‘행복’이 다소 높고 ‘흥미’는 매우 높게 나타났다. 또한, LG는 부정적 감성 중 ‘지루함’을 제외한 ‘슬픔’, ‘분노’, ‘혐오’, ‘공포’, ‘고

통’의 감성 비율이 모두 낮았다. 위의 분포를 토대로, 4대 기업 중에서 국민은 ‘LG’에 대해 가장 흥미를 가지고 있고 다소 긍정적인 감성을 가지고 있다고 할 수 있다. 반면에, ‘SK’에 대해서는 가장 지루함을 느끼고 있고 다소 부정적인 감성을 가지고 있다고 할 수 있다.

5.2 가설(1): 기업 간 감성 차이의 연구 결과

‘삼성’과 ‘한진’의 대응표본 T검정 결과를 토대로 ‘삼성’과 ‘한진’의 감성은 ‘공포’, ‘놀람’, ‘기쁨’, ‘흥미’, ‘고통’ 영역에서 유의적인 차이가 있다. ‘지루함’, ‘슬픔’, ‘분노’, ‘혐오’의 감성에서는 유의한 차이가 없었다(<Table 5> 참조).

비슷하게 부정적인 사건을 겪은 ‘SK’와 ‘한진’은 ‘놀라움’, ‘행복’, ‘지루함’의 감성에서 유의한 차이가 있었으며 나머지 영역의 감성에서는 유의한 차이가 없었다(<Table 6> 참조).

Duncan test의 결과, 기업은 ‘슬픔’ 감성이 낮은 순위대로 6개의 부분 집합으로 나뉘었으며 슬픔의 감성이 가장 낮은 것은 ‘LG’이고 가장 높

<Table 5> Paired samples test of Samsung & Hanjin

	Samsung & Hanjin - Paired Differences						t	df	sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference					
				Lower	Upper				
pair 1	disgust	-.02779	.09015	.01803	-.06501	.00942	-1.541	24	.136
pair 2	fear	.02333	.04555	.00911	.00453	.04213	2.561	24	.017
pair 3	surprise	.04722	.03851	.00770	.03132	.06311	6.131	24	.000
pair 4	happiness	-.07550	.14498	.02900	-.13535	-.01565	-2.604	24	.016
pair 5	interest	.02826	.04392	.00878	.01013	.04639	3.217	24	.004
pair 6	pain	-.01466	.03399	.00680	-.02869	-.00063	-2.156	24	.041
pair 7	rage	.00876	.05842	.01168	-.01535	.03287	.750	24	.461
pair 8	sadness	.01150	.09257	.01851	-.02672	.04971	.621	24	.541
pair 9	boredom	-.00111	.04588	.00918	-.02005	.01783	-.121	24	.904

<Table 6> Paired samples test of SK & Hanjin

		SK & Hanjin - Paired Differences					t	df	sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
pair 1	disgust	.00339	.10539	.02107	-.04011	.04690	.161	24	.873
pair 2	fear	.00357	.04595	.00919	-.01539	.02254	.389	24	.701
pair 3	surprise	.01732	.03667	.00733	.00218	.03246	2.362	24	.027
pair 4	happiness	-.12533	.15779	.03155	-.19047	-.06019	-3.971	24	.001
pair 5	interest	.00716	.04101	.00820	-.00977	.02409	.873	24	.391
pair 6	pain	-.00227	.04504	.00900	-.02087	.01631	-.253	24	.803
pair 7	rage	-.00370	.05726	.01145	-.02733	.01993	-.323	24	.749
pair 8	sadness	.01220	.09947	.01989	-.02885	.05326	.613	24	.545
pair 9	boredom	.08765	.07215	.01443	.05786	.11743	6.073	24	.000

<Table 7> Duncan test of 'sadness'

Group	N	Subset for alpha = 0.05					
		1	2	3	4	5	6
LG	25	.20619					
Posco	25	.23648	.23648				
Hyundai Motors	25	.24841	.24841	.24841			
Samsung	25		.25990	.25990	.25990		
Hyundai Heavy Industries	25		.26061	.26061	.26061		
Lotte	25			.29397	.29397	.29397	
SK	25				.30131	.30131	.30131
Hanjin	25				.30500	.30500	.30500
Hanhwa	25					.33828	.33828
GS	25						.34536
Sig.		.069	.323	.058	.067	.065	.067

Means for groups in homogeneous subsets are displayed.

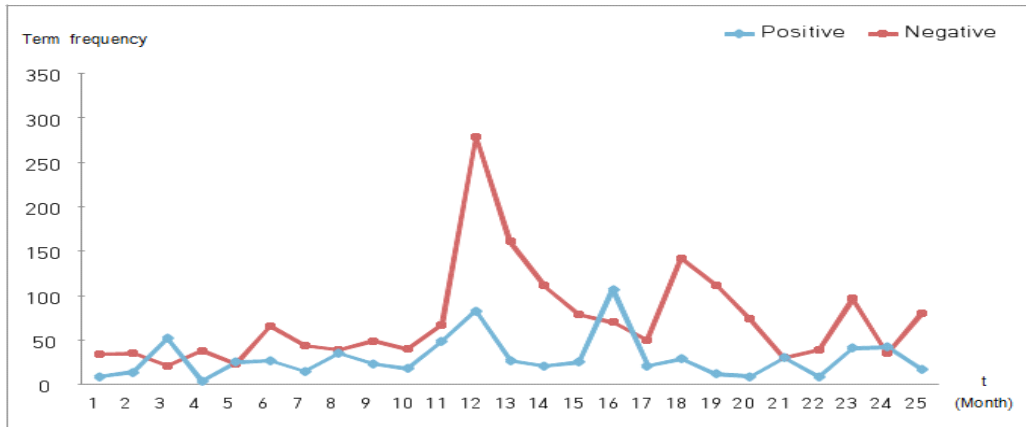
a. Uses Harmonic Mean Sample Size = 25.000.

은 것은 'GS'로 나타났다(<Table 7> 참조).

5.3 가설(2): 시점 간 감성 차이의 연구 결과

'땅콩회항'이 일어난 시점인 2014년 12월의 뉴

스 개수는 450개로 2014년 12월을 제외한 월 평균 뉴스 개수인 159.5개의 약 2.8배에 달한다. 감성 단어의 평균 검출 개수는 2014년 12월을 제외하고 월 평균 106개인 반면에 2014년 12월에 검



〈Figure 3〉 Changing frequency of positive & negative words of 'KOREANAIR'

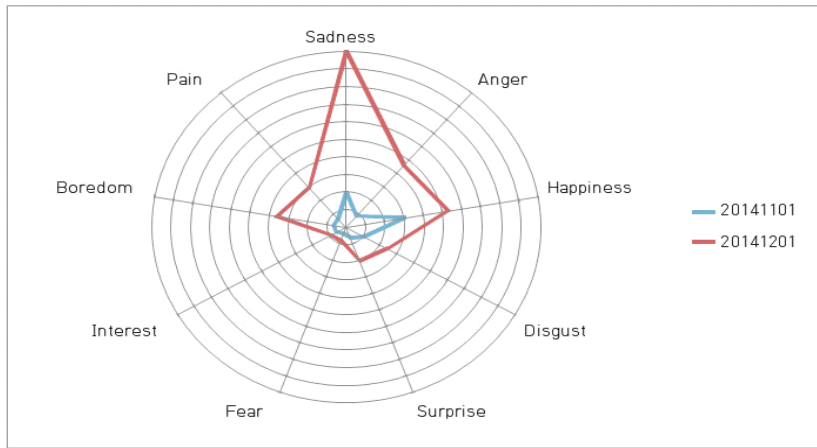
출된 감성 단어의 개수는 510개로 평균에 비해 약 4.8배 정도 많았다. ‘땅콩회항’ 사건은 국민의 감성에 영향을 끼친 사건이라고 할 수 있다.

긍정과 부정의 감성으로 본 ‘땅콩회항’ 사건은 국민에게 부정적인 영향을 크게 준 사건이라고 할 수 있다. 2014년 12월에 부정의 범주에 있는 감성 단어가 약 300개로 25개월의 표본 중 가장 많이 검출되었다(〈Figure 3〉 참조). ‘땅콩회항’ 사건의 영향력은 발생 이후 점차 감소하여 발생 이후 4개월이 경과된 시점에 긍정의 감성이 부정의 감성을 초월하였다. ‘대한항공’은 그 시점에 역사상 가장 높은 영업이익을 발표하였으며 기타 주가 상승과 각종 호재가 겹쳤기에 긍정이 부정을 초월한 것으로 볼 수 있다. 2015년 6월에는 다시 부정적인 감성 단어가 급증하였는데 그 이유는 ‘메르스’의 여파로 인하여 대한항공의 이용객 수가 크게 줄었고 그로 인하여 가치가 폭락했기 때문이다.

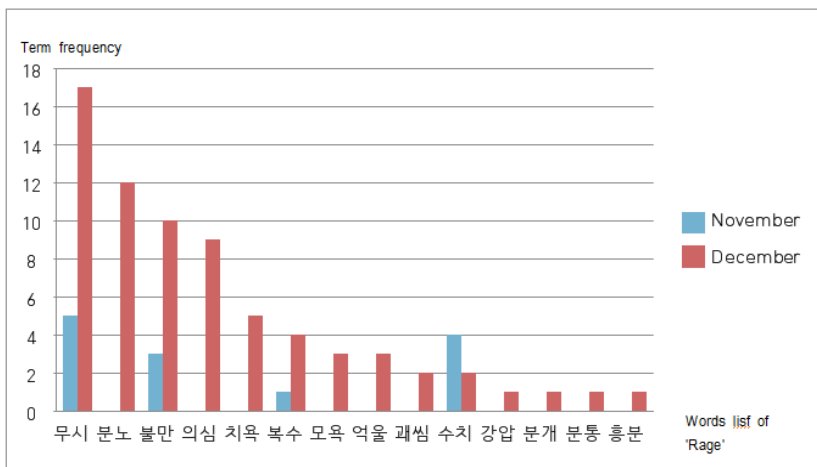
‘땅콩회항’ 사건이 일어난 2014년 12월과 사건 직전인 11월의 감성 단어 빈도 변화를 비교해보면 국민 감성 변화를 구체적으로 알 수 있다

(〈Figure 4〉 참조). 긍정에 해당하는 ‘행복’, ‘호기심’의 감성 단어 빈도보다 부정에 해당하는 ‘슬픔’, ‘분노’, ‘지루함’, ‘고통’의 감성 단어 빈도가 급격히 증가했다. 중립에 해당하는 ‘놀라움’의 감성 단어 빈도도 급격히 증가하였다. 반면에, 부정의 감성에 속하는 ‘두려움’과 ‘혐오’의 감성은 증가폭이 크지 않았다. ‘슬픔’, ‘분노’, ‘지루함’, ‘고통’의 감성이 다른 범주의 감성에 비해 유의한 차이를 보였기에 ‘땅콩회항’ 사건은 국민에게 위의 4가지 감성을 유발한 사건이라고 볼 수 있다.

4가지 감성 중에 ‘분노’에 해당하는 감성을 국민이 어떤 단어로 표현했는지 알고자 ‘땅콩회항’이 발생한 2014년 12월의 ‘분노’ 범주 단어 목록을 조사하였다(〈Figure 5〉 참조). ‘분노’ 범주에서 검출된 단어의 종류는 14개였고 가장 많이 검출된 단어는 ‘무시’로 17개가 검출되었다. 많이 검출된 순서대로 이어서 ‘분노’ 12개, ‘불만’ 10개, ‘의심’ 9개, ‘치욕’ 5개, ‘복수’ 4개, ‘모욕’ 3개, ‘억울’ 3개, ‘괘씸’ 2개, ‘수치’ 2개, ‘강압’ 1개, ‘분개’ 1개, ‘분통’ 1개, ‘홍분’ 1개가 검출되



〈Figure 4〉 Changing term frequency of 'KOREANAIR' (2014. 11. 01~2014. 12. 01)



〈Figure 5〉 Words list of 'Rage' in November and December, 2014

었다. '무시'는 매달 평균 검출 빈도가 1.6개임에도 불구하고 12월에만 17개가 검출되어 검출 빈도가 높게 증가한 단어라고 할 수 있다. 다른 달에는 검출되지 않았지만 12월에만 검출된 단어는 '분노', '치욕', '괘씸', '분개', '분통'으로 전체 검출된 14개 단어 중 5개의 단어가 '땅콩회항'이 발생한 12월에만 검출되었다. 사건이 발생하기

전 달인 2014년 11월과 발생한 달인 12월을 비교해보아도 그 차이를 알 수가 있다. 11월에는 '분노' 영역에서 단어가 '무시' 5개, '불만' 3개, '복수' 1개, '수치' 4개만 검출되었다. 검출된 단어 목록과 빈도의 차이를 바탕으로 국민의 감성을 추측할 수 있다. 분노의 감성 중에서도 국민은 구체적으로 14가지의 단어에 해당하는 분노의

감성을 느꼈다고 볼 수 있다.

6. 결론 및 시사점

6.1 연구 결과 정리

본 연구는 국내 10대 기업에 대한 뉴스 데이터를 대상 데이터로 하고, 개별 범주적 감성 구조를 기반으로 한 감성 사전을 활용하여 국민 감성 분석을 수행하였다. 개별 범주적 감성 구조는 특정한 분류 기준 없이 사람이 느낄 수 있는 감성을 보편적으로 정의하기에 빅 데이터 안에 있는 감성을 국민의 시각으로 측정하는데 효과적이라고 할 수 있다. 기존 선행 연구에서 주로 사용한 차원적 감성을 기반으로 한 감성 분석은 객관성이 있고 상관관계와 같은 각 차원의 관계를 알기가 쉽다는 장점이 있으나 직관성이 떨어져 국민의 보편적인 감성을 대변하기는 어렵다. 또한, 기준을 명확히 확립하는데 있어서 차원을 1차원이나 2차원으로 나누기에 복잡한 감성을 표현하는 데에 한계점이 있다. 개별 범주적 감성을 기반으로 한 감성 사전은 이러한 한계점을 보완한 하나의 대안이라고 할 수 있다. 감성을 9가지 범주로 나눈 선행 연구를 따라서 감성을 ‘슬픔’, ‘분노’, ‘기쁨’, ‘놀람’, ‘공포’, ‘혐오’, ‘지루함’, ‘흥미’, ‘통증’으로 분류하였다. 이는 기존에 주로 수행해왔던 긍/부정, 활성/비활성화와 같은 차원 기반의 감성 분석과는 차이가 있으며 국민의 관점에서 느끼는 감성에 집중했다고 볼 수 있다.

각 범주의 감성은 감성 단어의 검출 비율과 빈도를 기반으로 측정하였다. 국민은 기업 간 다른 감성을 가지며, 특정 사건은 국민의 감성에 영향

을 미칠 것이라는 가설을 세우고 검증하였다. 위의 가설은 개별 범주적 감성을 기반으로 한 감성 사전이 감성 분석에 있어 기업별·시점별 비교에 효과가 있는지 확인하기 위해 제시되었다.

검정 결과, 국민의 감성은 기업에 따라 차이가 있었다. 대응 표본 T검정을 수행하여 ‘삼성’과 ‘한진’, ‘SK’와 ‘한진’ 두 쌍을 비교하였고 기업 간 감성에 차이가 있음을 확인했다. 더불어, ‘슬픔’ 감성에 Duncan 검정을 수행하여 기업에 대한 감성은 순위적으로도 유의한 차이가 있음을 밝혔다. 이외에도 기업 간 감성을 비교하고자 각 기업에 감성 순위를 매겼으며 4대 기업은 따로 비교하였다. 4대 기업 중 ‘삼성’과 ‘현대자동차’의 감성은 큰 차이가 없었는데 반해, ‘SK’는 상대적으로 긍정의 감성이 낮고 부정의 감성이 높았으며 ‘LG’는 반대로 긍정의 감성이 더 높고 부정의 감성이 낮았다.

특정 사건은 국민의 감성에 영향을 주었다. ‘대한항공’의 ‘땅콩회항’ 사건을 통해 시점별로 감성을 비교하였다. 감성에 영향을 주는 사건이 있는 달에는 감성 단어의 검출 빈도가 비약적으로 증가했다. 뿐만 아니라, 9개의 감성을 긍/부정으로 다시 나누어 특정 사건의 긍정/부정적 영향을 확인할 수 있었다. 사건 전후 감성 변화를 전 범주에서 확인하여 사건이 어떤 감성에 영향을 주는지 알 수 있었다. 또한, 검출된 단어 목록의 변화를 토대로 국민의 구체적인 감성 변화 추이도 추정할 수 있었다.

개별 범주적 감성을 기반으로 구축한 감성 사전은 기업별·시점별 감성분석에 효과가 있다고 할 수 있다. 기존의 연구를 대체하는 것이 아니라 상호 보완하는 점에서 감성 연구는 더 정교해질 수 있을 것이라고 생각한다.

우리나라에서 대기업의 역할은 중요하다. 신

자유주의의 물결로 정부가 작아지고 기업의 역할이 커지고 있는 상황에서 향후 우리나라 10대 기업은 더 큰 역할을 행사할 가능성이 있다. 그러나 큰 역할을 행사한다고 해서 기업이 영원히 생존하는 것은 아니다. 통신 매체의 발달로 고객은 작게는 몇몇의 고객으로 결집하게도 하고, 많게는 국민으로 결집하여 집단적인 감성을 갖는다. 집단적인 감성의 파급효과는 빠르고 강하기에 기업은 감성을 측정할 수 있는 다양한 채널을 통하여 고객과 소통할 수 있어야 하며 나아가서는 미래의 감성도 예측할 수 있어야 한다. 요약하자면, 개별 범주적인 감성으로 기업에 대한 감성을 분석하는 것은 기업이 국민의 감성을 이해하는 측면에서 의미가 있다고 할 수 있다.

6.2 연구 한계점 및 향후 연구 방안

본 연구는 개별 범주적 감성을 기반으로 국민의 감성을 기업별·시점별로 측정할 수 있다는데 의의가 있다. 빅 데이터에서 감성을 측정하는 과정에서 오차가 발생할 수 있는 몇몇 한계점이 존재한다. 동시에, 이 측정 방식을 이용하여 감성의 움직임에 대응을 했을 때의 효과는 아직 증명되지 않았다. 본 연구가 갖는 한계점과 추후 연구 방안은 다음과 같다.

첫째로, 개별 범주적 감성 분류 방식이 가지고 있는 태생적인 한계점인 주관성이다. 감성 단어 사전은 선행 연구의 단어 목록에 의존하여 그 한계점을 그대로 가지고 있다. 감성을 분류하는 방식은 주관적일 수 있다. 학술적으로 정확히 ‘분노’와 ‘고통’의 영역을 나누는 것은 불가능하다고 할 수 있다. 설문 대상의 주관적인 판단이 개입되어 감성 단어가 결정되었을 가능성이 있다.

둘째로, 명사만 추출하거나 명사 형태로 단어

를 만들었기 때문에 감성을 측정하는데 오차가 있을 수 있다. 더불어, ‘R 패키지’의 ‘KoNLP’는 업데이트를 지속적으로 하고 있지만 기능이 완전하지 않아 오차가 생겼을 가능성이 있다.

셋째로, 문맥에 따른 단어의 진정한 의미는 파악하지 못했다. 같은 단어라도 맥락과 상황에 따라서 그 의미가 다를 수 있고, 말하는 주체에 따라서도 의미가 다를 수 있다. 이 점을 고려하지 못하고 빈도로 측정했기 때문에 오차가 발생할 수 있다.

위의 한계점은 본 연구가 가지고 있는 태생적인 한계점이라고도 할 수 있으며, 향후 연구에서 보완을 하더라도 오차를 줄일 수는 있지만 근본적인 한계점은 극복하기 어렵다고 볼 수 있다. 향후 연구에서 극복할 수 있는 한계점으로는 이어서 다음과 같다.

넷째로, 각 기업 그룹의 대표 계열사를 조사하여 전체 기업을 분석했기 때문에 기업에 대한 감성의 오차가 존재할 수 있다. 예를 들어서 같은 기업 집단에 속해 있다하더라도 ‘현대자동차’와 ‘모비스’에 대한 감성이 다를 수 있고 ‘LG화학’과 ‘LG전자’에 대한 감성이 같은 회사라도 다를 수 있다. 모든 사회사의 데이터는 감안하지 못했기에 향후 연구에서는 더 넓은 범위의 데이터를 포함시켜 오차를 줄일 필요가 있다.

다섯 번째로, 기업별 사전 구축이다. 뉴스는 표준어를 구사하고 문법을 정확히 준수하기에 통합 감성 사전으로 분석을 수행하기가 용이하다. 그러나 소셜 미디어에서 발생하는 비정형 데이터에도 본 연구의 감성 사전을 활용하기 위해서는 조절이 필요하다. 기업을 부르는 특정 은어나 비속어의 변형된 형태는 계속 출현하고 있기에 기업별로 따로 단어 사전을 추가할 필요가 있다.

마지막으로, 본 연구에서 사용한 측정 모형을 응용하여 실효성을 확인하는 것이다. 개별 범주적 감성을 이용하여 기업이 감성 측정·모형 적용·미래 예측을 할 수 있고 모형이 발전해나간다면 본 연구의 의의는 실용성 측면에서 확장될 수 있다. 예를 들어, 국민의 보편적 감성에 따른 주가 변화 예측이나 판매량 변화 예측과 같은 실효성을 증명할 수 있는 모형이 필요하다. 모형을 만드는 과정에서 정성적인 요소를 해석하여 정량적인 분석을 더 정교하게 만드는 방법이 필요하다. 가령, ‘대한항공’의 사례에서 그동안 검출되지 않았던 5개의 새로운 단어가 검출되었고 이 단어는 사건에 대한 감성을 설명하는 중요한 단어라고 할 수 있다. 빈도뿐만 아니라 감성을 파악하면 이 단어에 대한 중요성이 줄어들기에 설명력이 있는 단어에 대해서는 더 큰 가중치를 줄 필요가 있다. 이러한 방식은 감성을 더 구체적으로 측정하고 예측하는데 필수적인 요소라고 할 수 있다.

향후 위와 같은 한계점을 보완한다면 빅 데이터 안에 존재하는 국민의 감성을 더 정확하게 측정하고 기업 현장에서 효과적으로 사용하는 것이 가능해질 것이다.

참고문헌(References)

- Ahn, E. J. and Y. H. Hwang, “Theory and Practice of Lemma List Construction for a Dictionary – Focused on Yonsei Contemporary Korean Dictionary Compilation,” *Journal of Korealex*, Vol.15(2010), 165~193.
- Ahn, J. G. and H. W. Kim, “Building a Korean Sentiment Dictionary and Applications of Natural Language Processing,” *Proceeding of Journal of Intelligence and Information Systems*, Vol.2014, No.11, 177~182.
- An, J. Y., J. H. Bae, N. G. Han and M. Song, “A Study of ‘Emotion Trigger’ by Text Mining Techniques,” *Journal of Intelligence and Information Systems*, Vol.21, No.2(2015), 69~92.
- Ahn, S. H., S. H. Lee and O. S. Kwon, “A Study of Activation dimension: A mirage in the affective space,” *Korean Journal of Social Psychology*, Vol.7, No.1(1993), 107~123.
- Baek, B. H., L. K. Ha and B. C. Ahn, “An Extration Method of Sentiment Information from Unstructured Big Data on SNS,” *Journal of Korea Multimedia Society*, Vol.17, No.6(2014), 671~680.
- Cha, Y. S., J. H. Park, J. H. Kim, S. Y. Kim, D. K. Kim and M. C. Whang, “Validity analysis of the social emotion model based on relation types in SNS,” *Science of Emotion and Sensibility*, Vol.15, No.2(2012), 283~296.
- Choi, S. J., Y. E. Song and O. B. Kwon, “Analyzing Contextual Polarity of Unstructured Data for Measuring Subjective Well-Being,” *Journal of Intelligence and Information Systems*, Vol.22, No.1(2015), 83~105.
- Ekman, P. and H. Oster, “Facial Expressions of Emotion,” *Annual Review of Psychology*, Vol.30(1979), 527~554.
- Greenwald, M. K., E. W. Cook and P. J. Lang, “Affective judgment and psychophysiological response: Dimensional covariation in the evaluation of pictorial stimuli,” *Journal of Psychophysiology*, Vol.3, No.1(2007), 17~25.
- Jang, P. S., “Study on Principal Sentiment Analysis of Social Data,” *Journal of Korean Institute of Information Technology*, Vol.19,

- No.12(2014), 49~56.
- Jung, J. S, D. S. Kim and J. W. Kim, "Influence analysis of Internet buzz to corporate performance: Individual stock price prediction using sentiment analysis of online news," *Journal of Intelligence and Information Systems*, Vol.21, No.4(2015), 37~51.
- Kang, S. A., 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(2015), 1167~1173.
- Kim, D. H., T. M. Cho and J. H. Lee, "A Domain Adaptive Sentiment Dictionary Construction Method for Domain Sentiment Analysis," *Proceedings of the Korean Society of Computer Information Conference*, Vol.23, No.1(2015), 15~18.
- Kim, M. K., J. H. Kim, M. H. Cha and S. H. Chae, "An Emotion Scanning System on Text Documents," *Korean Journal of the Science of Emotion and Sensibility*, Vol.12, No.4(2009), 433~442.
- Kim, S. W. and N. G. Kim, "A Study on the Effect of Using Sentiment Lexicon in Opinion Classification," *Journal of Intelligence and Information Systems*, Vol.20, No.1(2014), 121~128.
- Kim, Y. S., N. G. Kim and S. R. Jung, "Stock-Index Invest Model Using News Big Data Opinion Mining," *Journal of Intelligence and Information Systems*, Vol.18, No.2(2012), 143~156.
- Kotler, P., "*Marketing 3.0: From Products to Customers to the Human Spirit*," 1000, Wiley, 2010.
- Kwon, O. K. and J. Heo, "Automatic Clustering of Korean Sentiment Words Based on Newspaper Articles," *Proceeding of Korean Information Science Society*, Vol.2014, No.12(2014), 147~149.
- Lee, D. H., H. K. Kang, S. H. Kim and C. M. Lee, "Autocorrelation Analysis of the Sentiment with Stock Information Appearing on Big-Data" *Korean Journal of Finance Engineering*, Vol.12, No.2(2013), 79~96.
- Lee, H. N., G. Y. Choi, S. W. Jung, S. J. Park and Y. S. Jung, "Strategic feeling defined through weight analysis of representative feelings," *Proceeding of Ergonomics Society of Korea*, 281~285.
- Lee, J. W., H. J. Song, E. K. Nah and H. S. Kim, "Classification of Emotion Terms in Korean," *Korean Journal of Journalism & Communication Studies*, Vol.52, No.1(2008), 85~116.
- Lee, K. B., J. B. Baik and S. W. Lee, "Estimating a Pleasure-Displeasure Index of Word based on Word Similarity in SNS," *Journal of KIISE : Computing Practices and Letters*, Vol.20, No.3(2014), 159~164.
- Lee, S. H, J. Choi and J. W. Kim, "Sentiment analysis on movie review through building modified sentiment dictionary by movie genre," *Journal of Intelligence and Information Systems*, Vol.22, No.2(2016), 97~113.
- Lee, S. Y, J. S. Ham and I. J. Ko, "A Classification and Selection Method of Emotion Based on Classifying Emotion Terms by Users," *Korean Journal of the Science of Emotion and Sensibility*, Vol.19, No.1(2016), 39~49.
- Park, I. C, "Study on Brand Image Enhancement

- and Sensitivity Advertising,” *The Treatise on The Plastic Media*, Vol.18, No.2(2015), 127~132.
- Park, I. J. and K. H. Min, “Making a List of Korean Emotion Terms and Exploring Dimensions Underlying Them,” *The Korean journal of social and personality psychology*, Vol.19, No.1(2005), 109~129.
- Rhee, J. W., H. J. Song, E. K. Na and H. S. Kim, “Classification of Emotion Terms in Korean,” *Korean Journal of Journalism & Communication Studies*, Vol.52, No.2(2008), 85~116.
- Rhee, S. Y, J. S. Ham and L. J. Ko, “A Classification and Selection Method of Emotion Based on Classifying Emotion Terms by Users,” *Korean Journal of the Science of Emotion and Sensibility*, Vol.15, No.1(2012), 105~120.
- Seo, J. H, H. J. Jo and J. T. Choi, “Design for Opinion Dictionary of Emotion Applying Rules for Antonym of the Korean Grammar,” *Journal of Korean Institute of Information Technology*, Vol.13, No.2(2015), 109~117.
- Sohn, S. J, M. S. Park, J. E. Park and J. H. Sohn, “Korean Emotion Vocabulary: Extraction and Categorization of Feeling words,” *Science of Emotion and Sensibility*, Vol.15, No.1(2012), 105~120.
- Song, M. J., “Tracking on Attention to the Emotion and Sensibility and its Application at the Innovative Companies: Focused on Content Analysis of Annual Reports,” *Science of Emotion and Sensibility*, Vol.19, No.1(2016), 39~48.

Abstract

Public Sentiment Analysis of Korean Top-10 Companies: Big Data Approach Using Multi-categorical Sentiment Lexicon

Seo In Kim*·Dong Sung Kim*·Jong Woo Kim**

Recently, sentiment analysis using open Internet data is actively performed for various purposes. As online Internet communication channels become popular, companies try to capture public sentiment of them from online open information sources. This research is conducted for the purpose of analyzing public sentiment of Korean Top-10 companies using a multi-categorical sentiment lexicon. Whereas existing researches related to public sentiment measurement based on big data approach classify sentiment into dimensions, this research classifies public sentiment into multiple categories. Dimensional sentiment structure has been commonly applied in sentiment analysis of various applications, because it is academically proven, and has a clear advantage of capturing degree of sentiment and interrelation of each dimension. However, the dimensional structure is not effective when measuring public sentiment because human sentiment is too complex to be divided into few dimensions. In addition, special training is needed for ordinary people to express their feeling into dimensional structure. People do not divide their sentiment into dimensions, nor do they need psychological training when they feel. People would not express their feeling in the way of dimensional structure like positive/negative or active/passive; rather they express theirs in the way of categorical sentiment like sadness, rage, happiness and so on. That is, categorial approach of sentiment analysis is more natural than dimensional approach. Accordingly, this research suggests multi-categorical sentiment structure as an alternative way to measure social sentiment from the point of the public. Multi-categorical sentiment structure classifies sentiments following the way that ordinary people do although there are possibility to contain some subjectiveness. In this research, nine categories: 'Sadness', 'Anger', 'Happiness', 'Disgust', 'Surprise', 'Fear', 'Interest', 'Boredom' and 'Pain' are used as multi-categorical sentiment structure. To capture public sentiment of Korean Top-10 companies, Internet

* School of Business, Hanyang University

** Corresponding Author: Jong Woo Kim

School of Business, Hanyang University

222 Wangsimni-ro, Seongdong-gu, Seoul 133-791, Korea

Tel: +82-2-2220-1067, Fax: +82-2-2220-1169, E-mail: kjw@hanyang.ac.kr

news data of the companies are collected over the past 25 months from a representative Korean portal site. Based on the sentiment words extracted from previous researches, we have created a sentiment lexicon, and analyzed the frequency of the words coming up within the news data. The frequency of each sentiment category was calculated as a ratio out of the total sentiment words to make ranks of distributions. Sentiment comparison among top-4 companies, which are ‘Samsung’, ‘Hyundai’, ‘SK’, and ‘LG’, were separately visualized. As a next step, the research tested hypothesis to prove the usefulness of the multi-categorical sentiment lexicon. It tested how effective categorial sentiment can be used as relative comparison index in cross sectional and time series analysis. To test the effectiveness of the sentiment lexicon as cross sectional comparison index, pair-wise t-test and Duncan test were conducted. Two pairs of companies, ‘Samsung’ and ‘Hanjin’, ‘SK’ and ‘Hanjin’ were chosen to compare whether each categorial sentiment is significantly different in pair-wise t-test. Since category ‘Sadness’ has the largest vocabularies, it is chosen to figure out whether the subgroups of the companies are significantly different in Duncan test. It is proved that five sentiment categories of Samsung and Hanjin and four sentiment categories of SK and Hanjin are different significantly. In category ‘Sadness’, it has been figured out that there were six subgroups that are significantly different. To test the effectiveness of the sentiment lexicon as time series comparison index, ‘nut rage’ incident of Hanjin is selected as an example case. Term frequency of sentiment words of the month when the incident happened and term frequency of the one month before the event are compared. Sentiment categories was redivided into positive/negative sentiment, and it is tried to figure out whether the event actually has some negative impact on public sentiment of the company. The difference in each category was visualized, moreover the variation of word list of sentiment ‘Rage’ was shown to be more concrete. As a result, there was huge before-and-after difference of sentiment that ordinary people feel to the company.

Both hypotheses have turned out to be statistically significant, and therefore sentiment analysis in business area using multi-categorical sentiment lexicons has persuasive power. This research implies that categorial sentiment analysis can be used as an alternative method to supplement dimensional sentiment analysis when figuring out public sentiment in business environment.

Key Words : Sentiment Analysis, dimensional sentiment structure, categorial sentiment structure, Multi-categorical sentiment lexicon

Received : August 18, 2016 Revised : September 13, 2016 Accepted : September 24, 2016
Publication Type : Regular Paper Corresponding Author : Jong Woo Kim

저 자 소개



김서인

현재 한양대학교 경영대학 경영학부 학사과정 재학 중이며, 주요 연구 관심분야는 데이터마이닝 기법과 응용, 빅데이터, 오피니언 마이닝, 핀테크 등이다.



김동성

현재 한양대학교 일반대학원 경영학과 경영정보시스템 전공 박사과정에 재학 중이다. 협성대학교 경영정보학과에서 학사를 마쳤으며, 한양대학교 경영학과에서 석사학위를 취득하였다. 주요 연구 관심분야는 데이터마이닝 기법과 응용, 오피니언 마이닝, 사회 네트워크 분석 등이다.



김종우

현재 한양대학교 경영대학 경영학부 교수로 재직 중이다. 서울대학교 수학과에서 학사를 마쳤으며, 한국과학기술원에서 경영과학으로 석사학위를, 산업경영학으로 박사학위를 취득하였다. 주요 연구 관심분야는 데이터마이닝 기법과 응용, 오피니언 마이닝, 상품 추천기술, 지능형 정보시스템, 집단지성, 사회 네트워크 분석, 기계학습 등이다.