

텍스트마이닝 기법을 활용한 한국인의 행복과 불행 탐색연구

An Exploratory Study of Happiness and Unhappiness Among Koreans based on Text Mining Techniques

박상현*, 도강혁**, 김학영***, 박가은****, 윤진혁*****, 김경일***
서강대학교 컴퓨터공학과*, 아주대학교 IT심리학과**, 아주대학교 심리학과***,
가톨릭대학교 심리학과****, 아주대학교 기계공학과*****

Sanghyeon Park(twodude@naver.com)*, Kanghyuk Do(kanghyukdo@gmail.com)**,
Hakyeong Kim(cs7610@naver.com)***, Gaeun Park(purin531@naver.com)****,
Jinhyeok Yun(wlsgurjin@ajou.ac.kr)*****, Kyungil Kim(kyungilkim@ajou.ac.kr)***

요약

본 연구에서는 텍스트 마이닝 분석을 통해 한국 사회에서 행복과 불행이 갖는 의미를 탐색하였다. 자료수집 및 분석을 위하여 온라인 뉴스 포털에서 Word2Vec과 TF-IDF 방법을 사용하여 '행복' 및 '불행' 키워드와 유사한 단어를 추출했다. 또한 K-LIWC 사전을 사용하여 행복 및 불행과 연관된 단어들의 감성 속성에 대해 알아보았다. TF-IDF 분석 결과, 행복과 불행은 사회적 요인과 해당 년도의 사회적 이슈들과 각각 높은 관련성이 있는 것으로 관찰됐다. Word2Vec 분석에서는 '희망'이 6년 연속으로 행복과 유사성이 높은 단어로 나타났다. K-LIWC 분석에서 '돈재정적이슈', '학교', '의사소통'은 행복 및 불행과 모두 관련성이 높았다. 그밖에 '몸 상태와 증상'이 불행과 높은 관련성이 있는 범주로 나타났다. 이러한 결과를 바탕으로 본 연구의 의의, 제한점 및 후속연구에 대한 필요성을 논의하였다.

■ 중심어 : | 행복 | 불행 | 빅데이터 | 텍스트 마이닝 | K-LIWC |

Abstract

The purpose of this study is to explore the meaning of happiness and unhappiness in Korean society through text mining analysis. Similar words with keywords(happiness/unhappiness) from online news portal are extracted using Word2Vec and TF-IDF method. We also use the K-LIWC dictionary to perform the sentiment analysis of words associated with happiness and unhappiness. In TF-IDF analysis, happiness and unhappiness are highly related to social factors and social issues of the year. In Word2Vec analysis, 'Hope' has been similar with happiness for six years. In K-LIWC analysis, 'money/financial issues', 'school', 'communication' is highly related with happiness and unhappiness. In addition, 'physical condition and symptom' is highly related to unhappiness. Implications, limitations, and suggestions for future research are also discussed.

■ keyword : | Happiness | Unhappiness | Big Data | Text Mining | K-LIWC |

I. 서론

인종과 국가, 시대를 불문하고 대부분의 인간이 살아 가며 추구하는 것 중 하나가 행복이다. 특히, 현대 사회에 들어와 행복에 대한 관심은 더욱 증가하고 있다. 다양한 정책에 ‘행복’이라는 단어가 붙고, 일상생활 속에서도 “웰빙(well-being)”, “올로족”이라는 단어들이 등장하는 것은 이러한 추세를 반영한다. 이처럼 행복은 쉽게 접할 수 있는 단어지만 개개인마다 행복을 추구하는 방식이 다르며 행복에 대한 정의도 다양하기 때문에, ‘행복이 무엇인가?’에 대해 분명하고 구체적인 정의를 내리기는 어렵다.

선행 연구들에서 행복에 대해 정의하려는 노력을 한 바 있지만[1-21], 아직도 많은 한계점을 가지고 있다. 먼저, 행복에 대한 정의는 문화권에 따라 달라질 수 있기 때문에[2] 해당 문화권에 맞는 행복에 대한 정의를 세우는 것이 필요하다. 본 연구에서는 국내 뉴스 포털에서 행복 및 불행 단어를 추출함으로써 현재 한국 문화권에서 행복 및 불행이 어떤 맥락으로 사용되는지를 탐색하였다. 또한 행복의 의미가 복합적이라는 것을 고려하여, K-LIWC 심리단어 사전을 사용하여 외적, 내적, 사회문화적 요인과 관련된 행복 및 불행의 다양한 범주들을 고려하였다. 이를 통해서 현재의 한국사회에서 행복이 어떤 의미를 가지는지 통합적으로 탐색하고, 행복 및 불행과 관련 범주들을 다양한 측면으로 비교할 수 있는 관점을 제공하였다.

본 연구에서는 빅데이터를 구축하기 위해 2000-2017년 각 언론사의 헤드라인과 기사 본문에서 ‘행복’과 ‘불행’이 포함된 전후 문장을 수집하였다. 이를 바탕으로 빈도분석을 실시한 후, TF-IDF와 Word2Vec 분석을 통한 중요 단어와 유사 단어 추출이 이루어졌다. 마지막으로 LIWC를 한국 문화에 맞게 타당화한 K-LIWC(korean-LIWC) 심리단어 사전을 사용한 감성 분석을 진행하였다. 이를 통해 네이버 뉴스 포털에 등장하는 행복 및 불행 관련 단어들을 34개 범주로 분류하고, 그 의미를 분석하였다. 이러한 빅데이터 분석 과정을 통해 기존의 설문지 연구에서 하지 못했던 대규모의 데이터 수집 및 분석이 이루어졌다. 또한 한국인들의

의 행복 개념의 변천사를 비교함으로써, 현재 한국 사회에서 행복이 가지는 의미를 체계적으로 정립하고 추후 연구를 위한 방향성을 제시하였다.

II. 관련 연구

다양한 시대에 걸쳐 많은 분야의 학자들이 행복에 대해 정의하고 논의하여 왔다. 특히, 철학 분야에서 행복에 대한 논의는 고대 그리스까지 거슬러 올라간다. 아리스토텔레스는 행복을 자아 실현적 관점에서 논하였고[22], 에피쿠로스학파에서는 쾌락주의적 관점에서 행복을 논하였다. 이 외에도, 과학적인 관점에서 행복이 연구된 것은 인간의 행동과 심리적 과정을 연구하는 심리학 분야 내에서였다.

행복 심리학이라는 비교적 최신 분야의 연구자인 Diener[1]는 행복 경험의 주관성을 강조하였으며, 행복을 개인이 자신의 삶에 대해 어떻게 느끼고 생각하는지와 관련된다고 주장했다. 이러한 맥락에서 그는 행복을 주관적 안녕감(subjective well-being)으로 정의하였으며 다음의 세 가지 특징을 언급했다. 첫째로, 주관적 안녕감은 ‘주관적(subjective)’이다. 즉, 개인의 내적 경험이라는 것이다. 건강이나 부, 평안함 등의 외적 요소는 행복에 영향을 줄 수 있지만 행복하기 위해 본질적으로 필요한 것은 아니다. 두 번째로, 주관적 안녕감은 긍정적인 측면을 반영하는 것으로 부정적인 측면과는 독립적이다. 즉, 부정적 정서의 부재만을 의미하는 것이 아니라 이와는 별개로 긍정적인 정서를 빈번하게 경험하는 것이 중요하다는 의미이다. 세 번째로, 주관적 안녕감은 개인이 삶 전체에 내리는 전반적인 평가라는 것이다. 이런 특징을 가진 주관적 안녕감은 크게 세 가지 요소로 개념화할 수 있다[1-3]. 첫 번째는 인지적 영역으로, 개인의 삶에 대한 전반적인 평가를 의미하는 삶의 만족도(life satisfaction)이다. 두 번째와 세 번째는 정서적 영역으로 긍정적 정서(positive affect) 경험의 빈도와 부정적 정서(negative affect) 경험의 빈도이다. 따라서 행복, 즉, 높은 주관적 안녕감은 긍정 정서를 많이 경험하면서 부정 정서를 적게 경험하고, 자신에 대해

긍정적으로 평가하는 것을 의미한다.

이후 Ryff[4]는 긍정 정서를 많이 경험하고 삶에 만족하는 것이 행복한 정서는 설명할 수 있지만 전 생애를 아우르는 진정한 행복을 설명하지 못한다고 비판하였다. 행복을 개인의 주관적 안녕감으로 정의한 Diener와 달리 Ryff[4]는 사회구성원으로서의 기능을 강조하며 행복을 '심리적 안녕감(psychological well-being)'으로 정의하였고 개인이 성격 강점, 덕목을 발현시켜 유능한 개체로 변형된 삶을 누리는 것을 의미한다고 언급하였다. 그는 진정한 행복을 경험하기 위해서는 개인이 자신의 내면적 능력을 개발하고 활용하여 삶의 질을 높여려는 노력을 하는 것이 필요하다고 주장하였다. 이에 따라 행복을 심리적 안녕감으로 정의하고, 자신에 대한 긍정적 태도(자기수용), 새로운 경험에 대한 개방성(개인적 성장), 삶의 목표를 갖는 것, 외부 환경에 대한 통제감(환경통제), 자기 주도적으로 일할 수 있는 능력(자율성), 타인과의 긍정적 인간관계 등 다섯 가지 요인을 행복의 구성 요인으로 제안하였다.

Keyes[5]는 Diener의 정서적 안녕감과 Ryff의 심리적 안녕감 개념이 너무 개인적 측면에만 치중되어 있다고 비판하고, 진정한 행복을 경험하기 위해서는 사회에 대한 긍정적 인식이 필요하다고 제안하였다. 이에 따라 행복을 사회적 안녕감으로 정의하고, 사회에 대한 긍정적 태도와 수용(사회적 수용), 타인에 대한 신뢰(사회적 실현), 사회에 대한 개인의 기여(사회적 기여), 사회의 잠재력에 대한 긍정적 평가(사회적 실현), 공동체에 대한 소속감(사회적 통합) 등의 다섯 가지 요인을 고려했다. 그밖에 최근 Keyes와 Magyar-Moe[6]는 주관적, 심리적, 사회적 안녕감이 행복의 서로 다른 측면을 반영한다고 제안하였으며, 정서적, 심리적, 사회적 요인과 그 하위요인들이 행복을 포괄적으로 정의한다고 제안했다.

2000년대 초, Martin Seligman이 인간의 강점, 가치, 안녕감 등 긍정적인 측면에 대한 연구의 필요성을 제기하며 시작된 긍정심리학에서도 행복은 중요한 연구 주제 중 하나였다. Seligman[7][8]은 행복을 긍정적인 정서, 만족 등 개인이 주관적으로 경험하는 좋은 감정이라 정의하였다. 행복은 세 가지 요소로 구성되어 있는

데, 긍정적인 정서(positive emotion)와 자신의 일, 관계, 여가 등에 열중하는 것(engagement), 그리고 의미(meaning)이다. 그에 따르면 행복한 삶이란 즐거운 삶, 몰입하는 삶, 의미 있는 삶이 조화를 이루는 것이다[7][8].

국내에서는 주로 설문지를 통해 개인의 주관적인 보고로 행복을 측정한 연구들이 주를 이루었다. 특히, 한국인의 행복 요인을 분석한 연구들이 많았다[9-12]. 또한, 스트레스[13][14], 사회적지지[15-17], 자아존중감[18][19], 성격 요인[20][21] 등 행복과 관련된 다양한 개인 내적, 외적 변인들에 대한 연구도 많이 수행되어 왔다.

이와 같이 행복에 대한 관심과 연구가 많이 이루어졌으나[1][4][5], 특히 최근 행복에 대한 관심이 폭발적인 이유는 경제적인 풍요에 비해 정신적 빈곤이나 박탈감을 느끼는 사람의 수가 증가하고 있기 때문일 것이다. 대표적인 예로, 대한민국의 경우 GDP 수준이 세계 11위에 도달한 국가이지만, 국가 행복지수에서는 10점 만점에 3.5점을 얻어 36개국 중 27위를 기록하는 데 머물렀다[23].

실제로 기존 연구 결과, 물질적인 부를 포함한 인구학적 변인들이 행복에 미치는 영향은 그다지 크지 않다는 것이 밝혀졌다. 연령, 성별, 수입, 인종, 교육, 결혼 상태를 포함한 인구통계학적 변인들은 행복의 개인차를 단지 8-15% 정도 밖에 설명하지 못한다고 보고된다[1][2]. 또한 행복과 다양한 내적, 외적 변인 간의 관계를 살펴 본 구재선과 서은국[24]의 연구에 따르면, 국내에서도 개인이 행복한 삶을 살아가는데 외적, 상황적 요인보다 개인의 심리적 특성이 중요하다는 사실이 확인되었다. 그러나 이러한 연구 결과들에도 불구하고 외적, 객관적 요인은 무시하기 어렵다. 행복에는 개인적 특성은 물론이고, 행복과 관련된 환경적 요인이나 활동도 영향을 미치며[25], 특히 행복과 관련된 이런 요인들은 문화에 따라 다르기 때문이다[2].

실제로 한국인들은 삶의 외적 조건이 행복을 결정하는 중요한 요소라는 강한 신념을 가지고 있는 것으로 나타났다. 2010년 서울시민 1,024명을 조사한 결과, 응답자의 40.5%가 행복을 위해서 돈이 필요하다고 응답했다[26]. 구재선 등[27]의 연구에서도 한국 대학 신입

생들이 미국 대학 신입생들보다 외적 가치를 중요시 여기는 경향이 더 높았고, 내적 가치를 중요시 여기는 경향은 더 낮았다. 박정현과 서은국[28]의 연구에서도 한국 대학생들은 돈, 집안배경 등의 외적인 요소를 행복의 조건으로 중요하게 생각하는 것으로 나타났다. 박영신, 김의철[29]의 연구에서는 경제적 측면이 삶의 질에 있어 가장 중요한 요소로 나타나기도 하였다. 이처럼 객관적인 외적 증거에 의해 자신을 평가하는 경향성이 행복 수준을 낮춘다는 다양한 선행 연구들을 고려할 때 [30][31], 삶의 외적 조건을 강조하는 한국 문화가 사람들의 행복 수준을 낮추는 중요한 요인일 수 있음을 짐작해 볼 수 있다.

또한 한국과 같은 집단주의 혹은 관계주의 문화권에서는 가족, 애인, 이웃과 같은 사회적 속성이 행복을 결정하는 중요한 요소일 수 있다. 박영신 등[32]의 연구에서는 경제적, 심리적, 관계적 자원이 어떻게 결합되어 주관적 행복감을 결정하는지 알아보았는데, 결과적으로 한국에서는 관계적 자원인 정서적 지지의 효과가 소득 효과에 비해 네 배 이상 큰 것으로 나타났다. 또한 한준[33]의 연구에서는 한국종합사회조사 자료를 토대로 주관적 행복감을 조사하였으며, 소득이 높은 사람과 자신을 비교할 때, 그리고 주변 관계에서 무시 받은 경험이 많을수록 자신을 불행하다고 생각하는 경향성이 있음을 밝혔다.

이와 같이 행복과 관련된 다양한 요인들이 한국문화 특정적으로 나타난다는 결과를 고려할 때, 한국 사회에서 행복이 어떠한 의미를 가지는지에 대해 재검증할 필요성이 있다. 아직까지 국내 연구에서는 외적 요인이나 사회적 속성보다 자기존중감, 낙관주의와 같은 심리적 요인과 행복의 관계를 밝히는 연구들이 주로 이루어져 왔으며, 외적 요인과 내적 요인, 사회적 속성을 함께 살펴본 연구는 드물다. 따라서 본 연구에서는 행복과 관련된 기존 연구들에서 개별적으로 관찰된 변인들을 통합적이면서도 다면적으로 살펴봄으로써, 한국인에게서 행복이라는 개념이 어떻게 정의되고 있는지 분명하게 규명하고자 한다.

이를 위해 본 연구에서는 빅데이터를 활용한 언어 패턴 분석을 활용하려고 한다. 언어 패턴 연구는 기존 사

회과학에서 자주 실시되었던 설문 연구를 보완할 수 있다. 설문지 연구는 응답 대상과 검사 시점, 검사 상황에 영향을 많이 받는다는 한계점을 갖기 때문이다. 반면 빅데이터는 표집편향의 문제를 개선해주며[34], 원하는 대상과 시점을 자유롭게 온라인상에서 선택하여 데이터를 추출할 수 있기 때문에 시공간적 제약에서 자유롭다. 본 연구는 2000-2017년 네이버 뉴스 포털에 실린 각 언론사의 헤드라인과 기사 본문에서 ‘행복’과 ‘불행’이 포함된 전후 문장을 추출함으로써, 빅데이터 분석을 위한 자료를 마련하였다.

또한 본 연구는 언어 분석 과정에서 단어들이 갖는 심리적 의미를 파악하기 위해 언어분석 프로그램인 LIWC(‘Linguistic Inquiry and Word Count’)의 심리적 범주와 단어 목록을 활용하였다. LIWC는 사람들이 작성한 글을 72개의 언어 관련 목록의 비율에 따라 분석하는 컴퓨터 프로그램으로, 우울과 같은 주요 심리학 연구에서 활용되어 왔다[35].

LIWC 프로그램은 빅데이터 관련 연구에서 주요 수단으로 활용되고 있다[34]. 대표적으로 페이스북과 같은 소셜 미디어 연구에서 LIWC 범주는 사용자들이 갖는 삶의 만족도를 예측할 수 있는 요소 중 하나이며 [36][37], 특히 정서 관련 범주는 사용자의 정서적 안녕감과 유의미한 상관이 있는 것으로 밝혀졌다[38]. 본 연구에서 사용된 K-LIWC(korean-LIWC) 심리단어 사전은 한국 문화에 맞게 타당화되었으며, 국내 연구에서도 주관적 삶의 질이 높은 집단과 낮은 집단을 비교하였을 때 K-LIWC 변인들에서의 유의미한 차이가 발견된 바 있다[39]. 따라서 본 연구는 K-LIWC 언어분석 프로그램을 한국인의 행복 및 불행을 탐색하기 위한 유용한 도구로써 사용하였다. 요약하면, 본 연구는 한국인의 행복 및 불행의 의미를 파악하기 위해, K-LIWC 언어분석 프로그램을 활용하여 국내 뉴스 포털에서 행복 및 불행과 관련된 단어들을 추출함으로써, 한국인의 행복 및 불행에 대한 의미를 탐색적으로 알아본 실증적 연구이다.

III. 연구 방법

3.1 데이터 수집

3.1.1 웹 크롤링

본 연구는 행복과 불행에 대한 사람들의 시간에 따른 인식 변화를 추적하기 위해 국내 뉴스를 웹 크롤링(web crawling)하여 데이터를 수집하였다. 수집을 위한 도구는 Python3을 사용하였고, 검색 키워드로는 ‘행복’과 ‘불행’이다. 데이터 수집 기간은 1990년 1월 1일부터 2017년 12월 31일까지이다. 그러나 포털에 뉴스가 본격적으로 제공된 시기가 2000년도이기 때문에[40], 실제 논문 분석에는 2000년부터 2017년까지의 데이터를 사용하였다. 수집 구간은 뉴스 헤드라인 및 본문 중 ‘행복’ 또는 ‘불행’ 키워드가 포함된 문장과 전후 문장 일부분을 발췌했다.

3.1.2 K-LIWC와 심리 단어 사전

K-LIWC는 LIWC를 원형으로 한국어의 특성과 문화를 반영하여 개발되었다[41]. K-LIWC는 크게 4개의 차원(언어적 구조, 심리적 내용, 자기영역[일상 관심사], 실험적 영역)을 통해 텍스트 내 언어와 단어 변인의 비율을 계산해주며, 심리학의 하위분야인 사회, 성격, 임상, 발달 심리에서 이론적으로 중요한 변인들을 측정하는데 효과적이어서 최근 널리 쓰이고 있다[42].

본 연구에서는 형태소 분석을 통해 얻게 될 ‘행복’ 또는 ‘불행’ 키워드와 연관된 단어들의 감성 속성을 알아보기 위해 K-LIWC의 한국어 심리 단어 사전을 활용하였다. 한국어 심리 단어 사전은 총 16,347개의 단어에 대해 총 44개의 심리학적 변인 속성을 갖는다. 사전에 포함된 단어 속성 범주의 관계를 분석한 결과, K-LIWC 심리 단어 사전은 34개의 독립적 속성과 10개의 종속성 속성으로 분석되는 것을 확인하였다. 예컨대 ‘긍정적인 느낌’과 ‘낙천성 또는 활동성’은 ‘긍정적인 정서’와 ‘감정 또는 정서적 과정’에 포함된다. 이에, K-LIWC에 속하는 단어에 대해서는 기존의 속성을 사용하고, 속하지 않는 단어에 대해서는 딥러닝 학습을 통해 단어들의 44개의 심리학적 변인 중 종속적인 10개 범주를 제외한 34개의 범주에 대한 포함 여부를 예측하였다.

3.2 형태소 분석

형태소 분석(morphological analysis)의 전체적인 과정은 [그림 1-a]에 요약되어 있다. 웹에서 수집된 원본 크롤링 데이터는 Python의 KoNLPy 패키지에서 제공하는 KOMORAN 형태소 분석기를 통해 ‘형태소 및 품사가 태깅된 단어 목록’을 생성한다. 실질적인 의미를 가진 ‘일반명사’와 ‘고유명사’에 해당하는 단어만을 추출해 말뭉치(corpus)를 형성해 분석에 사용한다. 이 말뭉치는 서로 다른 2가지 단어 임베딩(word embedding) 기법(TF-IDF, Word2vec)을 사용하여 각각의 연도별 ‘행복’ 및 ‘불행’ 키워드와의 연관 단어를 분석한다.

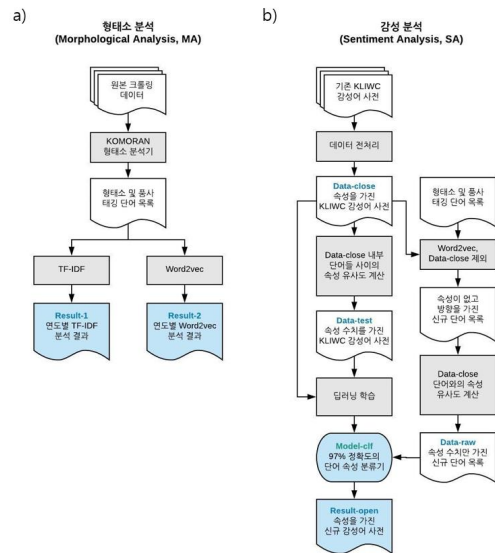


그림 1. 형태소 분석과 감성 분석 과정

3.2.1 KOMORAN 형태소 분석기

KOMORAN 형태소 분석기는 여러 어절을 하나의 품사로 분석 가능하며 공백이 포함된 고유명사를 더 정확하게 분석 가능하다는 특징을 가진다. 웹 크롤링을 통해 수집된 문장은 데이터 분석을 위해 다음과 같이 변환되었다.

- 일반명사 및 고유명사만 취한다.
- 중복 키워드는 제외한다.
- 불용어일 경우는 제외한다.

변환된 단어 목록은 TF-IDF 및 Word2vec 기법을 사용하여 분석되었다.

3.2.2 TF-IDF

TF-IDF(term frequency - inverse document frequency)[43][44]는 정보 검색과 텍스트 마이닝에서 이용하는 가장치로, 여러 문서로 이루어진 문서군이 있을 때, 어떤 단어가 특정 문서 내에서 얼마나 중요한 것 인지를 나타내는 통계적 수치이다. TF-IDF는 문서의 핵심어를 추출하는 용도로 주로 사용된다. TF(term frequency)는 특정한 단어가 개별 문서에서 얼마나 자주 등장하는지를 나타내는 값으로, 이 값이 높을 수록 개별 문서에서 중요하다고 생각할 수 있다. 하지만 해당 단어가 전체 문서 내에서 자주 사용되는 경우, 그 단어가 흔하게 등장한다는 것을 의미한다. 이것을 DF(document frequency)라고 하며, 이 값의 역수를 IDF(inverse document frequency)라고 한다. TF-IDF는 TF와 IDF를 곱하여 전체 문서군을 대상으로 어떤 단어가 특정 문서 내에서 얼마나 중요한지를 나타낼 수 있게 한다. 본 연구에서는 TF의 결과에 더 가장치를 두기 위해 IDF의 로그 값을 사용하여 형태소 및 품사 태깅 단어 목록에 TF-IDF를 수행하였다.

3.2.3 Word2vec

Word2vec는 단어가 가지는 의미를 다차원 공간에서 벡터(vector)화 하는 방식으로, 각 단어들 사이의 유사도를 측정할 수 있다. Word2vec는 단어의 의미가 공간상의 벡터로 표현되기 때문에, 벡터 연산을 통해 단어들 사이의 관계를 추론할 수 있다. [그림 2]는 이러한 Word2vec를 통한 단어의 의미 벡터 예시를 나타낸다 [45]. Word2vec가 잘 수행되면 [그림 2-a]와 같이 남자와 여자 사이의 벡터는 왕과 여왕 사이의 벡터, 삼촌과 이모 사이의 벡터 비슷하게 예측되는 것을 확인할 수 있다. 이를 활용하면 단어들 사이의 벡터 연산이 가능하다는 것을 [그림 2-b]로확인 가능하다. 이는 각 단어들의 코사인 유사도(cosine similarity)를 계산함으로써 이루어진다. 코사인 유사도는 다차원 공간상의 두 벡터간 방향의 유사도를 나타내는 수학적 정의이다. 이 값

은 두 벡터의 방향이 완전히 같을 경우 1의 값을 갖고, 반대 방향인 경우 -1의 값을 가지게 된다. 본 연구에서 수행된 Word2vec 공간에서 두 단어 사이의 코사인 유사도를 구하면 두 단어가 유사한 의미 관계에 있는지, 상반된 의미 관계에 있는지 알 수 있다.

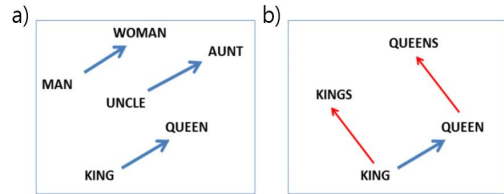


그림 2. Word2vec를 통한 단어 의미 벡터화 예시

3.3 감성 분석

감성 분석(sentiment analysis)의 전체적인 과정은 [그림 1-b]에 요약되어 있다. 감성 분석의 목적은 수집한 단어 목록의 심리학적 변인 속성을 알아보는 것이다. 이를 위해 기계 학습(machine learning)과 결합한 감성 분석(sentiment analysis)의 도구로서 자주 활용 [46]되는 LIWC의 한국판 버전인 K-LIWC를 적용하였다. 하지만 형태소 분석을 통해 분석한 다수의 단어는 K-LIWC 단어 사전에 속하지 않는 문제가 있어, 본 연구에서는 기존의 K-LIWC 단어 사전을 사용한 딥러닝 학습을 통해 새로운 단어들의 심리학적 변인 속성을 예측해 보았다.

3.3.1 K-LIWC 데이터 전처리

딥러닝(deep learning) 학습은 기계 학습의 한 분야로 최근 각광받고 있는 기술이다. 본 연구에서는 심층 신경망(Deep Neural Network, DNN) 구조를 사용하였다. DNN은 입력층(input layer)과 출력층(output layer) 사이에 여러 개의 은닉층(hidden layer)을 갖춘 인공신경망(Artificial Neural Network, ANN)이다. DNN 학습을 위해서는 학습을 위한 입력과 출력 데이터 쌍을 구성해야 한다. 학습은 이러한 입력과 출력 사이의 연관성을 찾아내도록 이루어진다. 기존 K-LIWC 단어 사전에서 [그림 1-a]에서 얻은 말뭉치에 속하면서 동시에 속성이 없는 단어를 제외하면 8,963개의 속성을 가진 K-LIWC

단어 목록을 얻을 수 있는데, 이를 Data-close라 하겠다. Data-close는 [그림 3]과 같이 구성되는데, 각각의 단어들의 44개 속성 해당 여부를 0과 1로 전처리하였다.

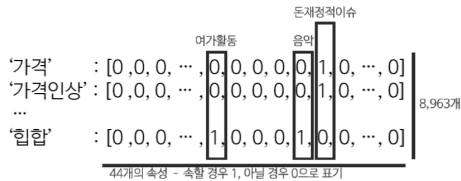


그림 3. Data-close 예시

3.3.2 단어들 사이의 속성 유사도 계산

[그림 1-a]에서 얻은 말뭉치에 속하면서 동시에 Data-close에 속하지 않는 단어 목록을 구하면 ‘속성이 없고, (Word2vec 공간에서의) 방향을 가진 신규 단어 목록’을 생성할 수 있다. 이 단어 목록에 속하는 신규 단어를 Data-close에 속한 8963개의 단어 각각에 대해 코사인 유사도를 구하고 각 속성 값(0 또는 1)을 곱해 속성별 합을 구하면 ‘속성 수치를 가지는 신규 단어 목록’을 생성할 수 있다. 이를 Data-raw라 하겠다. Data-raw는 [그림 4]와 같이 속성 수치를 가진다.

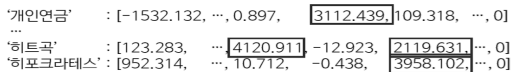


그림 4. Data-raw 예시

이 과정을 통해 신규 단어를 1개의 속성만 결정하기 위해서는 가장 큰 값만 취하면 된다. 하지만 이 방식을 취하면 [그림 4]에서처럼 ‘히트곡’의 경우처럼 여러 개의 속성이 존재하는 것처럼 보이는 경우를 판단하기 힘들다. 따라서 본 연구에서는 딥러닝을 통해 신규 단어의 속성 개수를 예측해 보았다. 정답이 있는 문제를 얻기 위해 Data-close도 Data-raw를 생성하는 방식과 유사한 방식으로 [그림 4]와 같이 ‘속성 수치를 가진 K-LIWC 단어 사전’을 생성하였다. 이를 Data-test라 한다.

3.3.3 딥러닝 학습

DNN 구조는 완전히 연결된 다층 신경망이다. 은닉층은 1개, 은닉 노드는 200개를 사용하였고, 손실 함수는 binary cross-entropy를 적용하였다. compile에는 손실 함수를 최소화하는 역전파(Backpropagation) 알고리즘[47]의 변형인 Adam optimizer를 사용하였다. 학습에 필요한 입출력 데이터는 각각 Data-test와 Data-close이다. 활성화 함수는 은닉층에 대해 ReLU 함수를 사용하였고, 출력층은 sigmoid 함수를 사용하였다. 과적합(overfitting)을 방지하기 위해 드롭아웃(Dropout) 기법이 적용되었다.

3.3.4 K-LIWC 심리 단어 사전 확장

[그림 5]는 딥러닝 학습을 통해 신규 단어의 속성을 예측하는 개략도를 나타낸다. 인공신경망 학습은 기존 K-LIWC 심리 단어 사전을 사용하여 이루어지고, 이렇게 학습된 인공신경망을 Model-clf라 하겠다. Model-clf는 약 97%의 정확도를 가지며, 이를 이용하여 Data-raw의 속성을 예측할 수 있다. 예측된 ‘속성을 가진 신규 심리 단어 사전’을 Result-open이라 하고, [그림 6]는 Result-open의 예시를 보여준다.

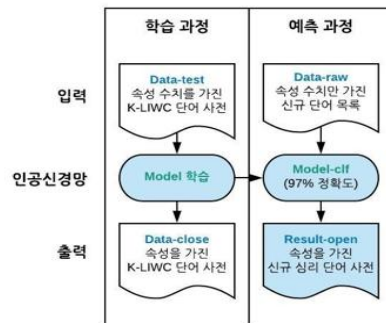


그림 5. 인공신경망 기반 속성 예측 개략도

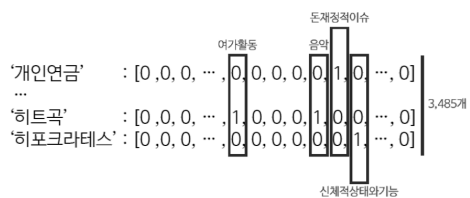


그림 6. result-open 예시

본 연구에서 속성을 예측한 Result-open의 개수는 총 3,485개로 구성되며, 각 단어들은 복수의 속성을 가진다. 실제 새로운 단어 사전 Result-open을 생성함에 있어 몇 가지 추가적인 고려사항이 있다. 첫 번째는 속성별 등장 빈도에 따른 가중치를 고려한 것이다. 두 번째는 44개의 속성 중 34개의 독립적 속성만으로 학습 수행하였고, 속성 예측 후 종속성을 복구하여 다시 44개의 속성으로 재분류하였다는 점이다. 따라서 새롭게 확장된 K-LIWC 사전인 Result-open 역시 기존 LIWC 사전과 동일하게 44개의 범주를 갖는다.

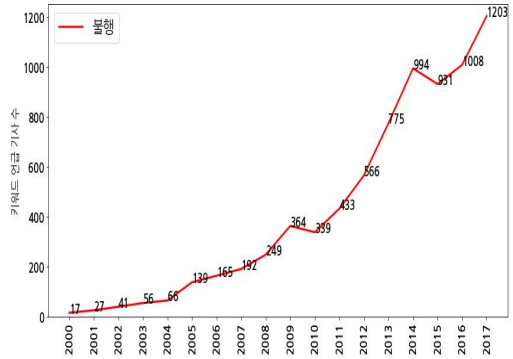


그림 8. 2000 ~ 2017년 간 불행 키워드 언급 기사량의 변화

IV. 연구 결과

4.1 기사 헤드라인 및 키워드 문장을 통한 행복 및 불행의 빈도 분석

한국인들의 행복과 불행을 이해하기 위해서는 한국 사회에서 행복과 불행이 어떻게 논의되어 왔는지 파악할 필요가 있다. 이를 위해 2000년부터 2017년까지 수집된 기사 데이터를 바탕으로 빈도분석을 진행하였다. ‘행복’은 총 405415 건, ‘불행’은 총 7565 건이 수집되었으며, 이를 바탕으로 한 연도별 행복 및 불행 단어 언급 횟수는 [그림 7][그림 8]과 같다.

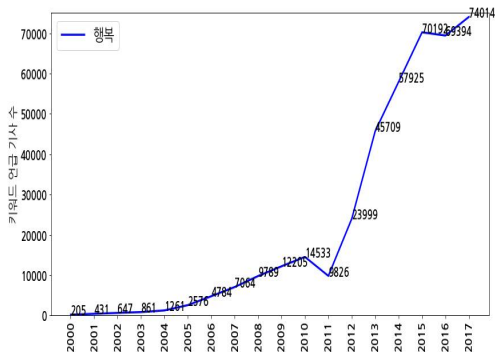


그림 7. 2000 ~ 2017년 간 행복 키워드 언급 기사량의 변화

분석 결과 불행 기사들의 중앙값은 294건, 행복 기사들의 중앙값은 9807.5건으로 행복에 대한 기사가 훨씬 많았다. 또한 행복과 불행에 대한 언급은 꾸준히 증가하는 추세로 나타났는데, 행복은 2000년 205건에서 2017년 74014건으로 약 361배 증가하였고, 불행은 2000년 17건에서 2017년 1203건으로 약 70배 증가하였다. 키워드 언급 기사의 빈도가 증가하는 것은 통신사와 종합편성채널 등 새로운 공급자들이 등장하면서 온라인 포털에 뉴스를 제공해왔기 때문이다[48]. 그럼에도 불구하고 주목할 만한 사실은 2011년 이후 행복을 언급하는 횟수의 증가폭이 매우 빠르다는 것이다. 2012년은 전년도에 비해 144% 증가하였으며, 2013년은 90% 증가하였다. 또한 불행 역시 2012년 30.7%, 2013년 36.9% 등 높은 증가세를 보였다.

본 연구에서 사용하는 Word2Vec과 같은 방법론은 학습되는 데이터의 양이 중요한 것으로 알려져 있다 [49]. 이러한 점을 고려하여, 앞으로의 분석에서는 한국 사회의 행복 및 불행 관련 논의가 본격화되어 충분한 데이터를 제공하기 시작한 2011년 이후에 초점을 맞추었다.

4.2 행복 및 불행 관련 신문기사의 연도별 중요도 분석

본 연구에서는 TF-IDF의 상위 20개 값을 표로 제시하였다. 행복 및 불행 기사의 연도별 TF-IDF 목록은 [표 1]과 [표 2]에 기술하였다.

표 1. 행복 관련 기사의 연도별 TF-IDF 주요 단어

연도	TF-IDF 값 범위	단어 목록
2011	164.6 ~ 669.7	이하정, 기태영, 정준호, 포도, 유진, 재단, 결혼식, 캠페인, 광장동, 봉사단, 포즈, 배우, 엔, 이벤트, 한태유, 이천희, 전해진, 팬, 아나운서, 미소
2012	387.0 ~ 1388.6	새누리당, 포도, 재단, 결혼식, 캠페인, 박근혜, 어르신, 김중인, 다이망, 고객, 미소, 봉사단, 포즈, 정준하, 엔, 센터, 복지, 협약, 다문화가정, 개최
2013	926.0 ~ 5315.0	기금, 주택, 채무, 포도, 한국자산관리공사, 결혼식, 국토교통부, 지구, 도경완, 접수, 이민경, 시범, 협약, 연세, 강남구, 선예, 신재윤, 금융위원회, 복지, 미소
2014	1081.1 ~ 2887.1	레드벨벳, 주택, 국토교통부, 시민, 지역, 박람회, 복지, 협약, 가오쓰치, 어르신, 농업, 소감, 사업, 봉사단, 포도, 이충재, 센터, 개최, 마을, 재단
2015	1450.3 ~ 3202.4	주택, 인스타그램, 국토교통부, 봉사단, 드림, 지역, 복면, 센터, 복지, 마을, 운영, 학습, 시민, 협약, 맥도날드, 개최, 이충재, 이벤트, 박람회, 선정
2016	1489.5 ~ 3673.4	주택, 인스타그램, 이충재, 홀씨, 국토교통부, 국제뉴스, 협약, 어르신, 지역, 센터, 스톱, 시민, 복지, 사업, 개최, 운영, 드림, 제공, 행정중심복합도시건설청, 마을
2017	1687.8 ~ 3995.3	인스타그램, 복지, 농업, 주택, 어르신, 농협, 국제뉴스, 센터, 지역, 사업, 이하율, 이윤지, 시민, 협약, 손승원, 개최, 운영, 교육지원청, 바우, 제공

행복 관련 기사의 TF-IDF 분석을 통해 행복 관련 연도별 주요 사건을 유추할 수 있다. 예컨대, 인명과 고유 명사를 제외한 단어들을 살펴보면, 2012년에서 새누리당이 언급된 것은 새누리당이 경선에서 과반을 넘는 승리를 거둔 사실을 반영하고 있다.

또한, 행복의 경우 ‘재단’, ‘센터’, ‘지역’, ‘복지’, ‘봉사단’ 등과 같이 사회적 요인과 관련된 단어들이 2011년도부터 2017년도까지 꾸준히 언급된 것을 확인할 수 있었다. 불행의 경우 해당 년도에 큰 이슈가 되었던 부정적 사건들을 반영하고 있었다. 2011년도의 경우 모 연예인의 ‘병역기피’, 2013년도는 ‘개성공단’ 중단, 2014년도는 ‘세월호’ 사건, 2016년의 경우 ‘새누리당’, ‘탄핵’, 2017년도의 경우 ‘박근혜’, ‘구치소’ 등의 단어가 불행과 관련이 높은 그 해의 특징적인 단어들로 나타났다.

표 2. 불행 관련 기사의 연도별 TF-IDF 주요 단어

연도	TF-IDF 값 범위	단어 목록
2011	10.9 ~ 21.7	계시, 대자보, 애니스톤, 해경, 병역기피, 조관우, 구청장, 라이덴, 란, 모래폭풍, 무소륵스키, 열꽃, 사임, 피트, 전성운, 브래드, 타블로, 조동달, 이병헌, 유순실
2012	17.3 ~ 69.5	정두언, 황상민, 새누리당, 액땀, 일본축구협회, 엘피다, 영합, 김연아, 중조, 경제민주화, 인명진, 민주통합당, 김경탁, 심재철, 안철수, 구니, 박근혜, 모로시니, 근대사, 닥터진
2013	21.7 ~ 69.3	채동욱, 계산사, 오지현, 고아성, 설국열차, 개성공단, 직장인, 페이스북, 이석기, 김찰총장, 이소라, 새누리당, 피물, 주택, 기금, 예산안, 이지가, 달빛프린스, 잔고, 통합진보당
2014	60.2 ~ 272.3	송영선, 세월호, 차승원, 이수진, 허지웅, 송남이, 이화여대, 전용우의, 시사집중, 침몰, 차노아, 중앙아프리카공화국, 학창, 부룬디, 마녀사냥, 해피엔딩, 정비, 차드, 콩고민주공화국, 등침
2015	27.3 ~ 189.4	김구라, 라디오스타, 성원중, 백야, 13일의 메르스, 대졸, 조강지처, 자영업, 비정상회담, 경매, 종사, 피습, 캡처, 검찰, 전문직, 상류사회, 토크, 압구정, 마녀
2016	29.4 ~ 83.4	필리버스터, 새누리당, 국민의당, 탄핵, 양극단, 이성민, 이용섭, 정진석, 최순, 개현, 홍윤, 시국, 이영범, 노유경, 캡처, 안철수, 탈당, 수업, 가화, 리멤버
2017	34.5 ~ 141.6	한국당, 탄핵, 바른, 박근혜, 서유기, 구치소, 구속, 파면, 헌정사, 국민의당, 한광옥, 정당, 블랙리스트, 주호영, 손나은, 문재인, 하태경, 김무성, 정우택, 개헌

TF-IDF는 비교 대상이 되는 두 문서에 모두 포함된 단어만이 값에 영향을 미치는 방법론이다[49]. 따라서 TF-IDF를 통해 문서군 내 핵심 단어를 선정할 수 있지만, 이 단어들은 주변 맥락까지 포함한 의미를 파악하기 힘들 수 있다. 이러한 한계점은 행복과 불행에서 동시에 나타나는 단어를 통해서 확인할 수 있다. 예컨대 ‘주택’이라는 단어는 행복 문서와 불행 문서 모두에서 발견된다. 이는 주택이 행복과 불행에서 중요하게 다뤄지는 사건(event)임을 의미한다. 그러나 주택이라는 단어 자체가 행복 혹은 불행 중 어느 한 개념과 보다 유사한지, 행복과 불행 각각에서 어떤 맥락을 통해 등장하는지를 판단하기는 어려울 것이다.

이 문제를 해결하기 위해 본 연구에서는 Word2Vec 기법을 사용하여 행복과 불행에 관련된 유사 단어를 직접적으로 도출하였다. 이를 통하여 행복과 불행과 관련된 유사 단어를 파악하고 맥락을 포함한 의미를 고려할 수 있다.

4.3 행복 및 불행 관련 신문기사의 연도별 유사도 분석

행복과 불행 기사에 대한 Word2Vec의 상위 20개 값을 표로 제시하였다. 행복 기사의 연도별 Word2Vec 목록은 [표 3]에, 불행 기사의 연도별 Word2Vec 목록은 [표 4]에 기술되었다.

표 3. 행복 관련 기사의 연도별 Word2Vec 주요 단어

연도	Word2Vec 값 범위	단어 목록
2011	0.79 ~ 0.85	일, 준비, 때, 불행, 아이, 삶, 부모, 지식, 영위, 은퇴, 양립, 인생, 책, 계, 생각, 자녀, 가전, 연구, 꿈, 생활
2012	0.51 ~ 0.64	송재호, 착안, 기원, 악기, 희망, 한부모, 어려움, 평등, 열기, 활력, 존중, 주방, 앞, 해피, 식생활, 원동력, 종이, 영종, 주기, 멘토링
2013	0.38 ~ 0.50	희망, 소생, 소명, 열린, 정감, 상상, 물, 용문면, 남수단, 모범, 윤경희, 병무청, 동시, 전기자동차, 가동, 벤치, 고리, 공염불, 협동, 사제
2014	0.34 ~ 0.47	희망, 선행, 푸른, 을, 수용자, 3포, 감동, 각각, 뜻, 법번, 출발, 민방위, 입가, 곳곳, 아지트, 아래, 보미, 기차, 풍토, 대청호
2015	0.33 ~ 0.52	희망, 을, 소방시설, 준비, 출발, 게스트하우스, 밀거름, 의미, 감동, 필요, 의지, 졸업말, 일반, 결실, 정기, 활기, 맛춤, 불당, 올해, 금연
2016	0.36 ~ 0.46	시범학교, 교육관, 희망, 역지사지, 계기, 일환, 특색, 열매, 소리, 발, 치유, 자리매김, 공정, 버팀목, 주제, 만전, 취지, 활력, 예정, 담벼락
2017	0.34 ~ 0.45	취지, 가동, 호응, 시동, 약속, 영제, 지역민, 우수지, 소망, 품앗이, 희망, 나무, 설명, 전인, 목표, 마음, 계기, 개보, 발, 민의

표 4. 불행 관련 기사의 연도별 Word2Vec 주요 단어

연도	Word2Vec 값 범위	단어 목록
2011	0.91 ~ 0.99	때, 생각, 말, 마음, 사람, 계, 인생, 아이, 일, 얼마, 지금, 이야기, 삶, 준비, 정답, 이유, 꿈, 일상, 노래, 후
2012	0.79 ~ 0.86	이유, 선택, 상황, 과거, 존재, 보통, 생각, 사실, 황지, 지안, 가난, 고민, 해민, 답, 여당, 침묵, 일치, 도전, 조건, 돈
2013	0.73 ~ 0.81	앵커, 멘트, 사실, 현실, 잘못, 예측, 반면, 권력, 수치, 답, 언급, 진실, 대담, 논리, 갈등, 잘못된, 과거, 사태, 돈, 정의
2014	0.66 ~ 0.72	이유, 잘못, 비교, 로또, 교통, 가난, 성향, 발언, 불안, 타인, 짧은이, 현명, 절망, 상처, 부유, 직면, 중앙아프리카공화국, 시절, 척도, 민주
2015	0.66 ~ 0.72	가난, 출세, 의사, 정의, 숫자, 절대, 복수, 수치, 고통, 차이, 총량, 권력, 근원, 불안, 반면, 분노, 확산, 주장, 인간, 대담
2016	0.67 ~ 0.75	정치, 분열, 권력, 사태, 정권, 존재, 위협, 대신, 다행, 추락, 생존, 진실, 존엄, 야당, 전체, 차이, 헌법, 잘못, 부탄, 집권
2017	0.68 ~ 0.78	반복, 가난, 호소, 희생, 성적순, 탓, 불안, 죽음, 주장, 권력, 존재, 짜지우지, 사건, 정치, 집착, 고통, 파괴, 다행, 퇴진, 중독

Word2Vec을 통해 행복 관련 연도별 유사 단어를 추출하였다. ‘희망’과 같은 단어는 2012년부터 2017년까지 6년 연속으로 행복과 유사성이 높은 단어로 나타났다.

반면, 불행 관련 Word2Vec 분석 결과 ‘가난’과 ‘생존’, ‘출세’와 ‘성적순’ 같은 단어들이 불행과 유사성이 높은 단어로 나타났다. 이는 불안한 미래가 한국 사회의 불행 관련 토픽으로 선정된 이연경 등[50]의 연구와 일치한다고 하겠다.

4.4 감성 사전을 통한 행복 및 불행 관련 기사의 K-LIWC 분석

TF-IDF 분석과 Word2Vec을 통해 행복 및 불행 기사에서의 핵심 단어와 유사성 높은 단어를 추출하였다. 본 연구는 이러한 분석을 심화하여 행복과 불행에 관련된 의미를 보다 다양한 측면으로 파악하기 위한 시도가 이루어졌다.

행복과 불행은 모두 추상적인 개념으로, 개인 및 문화에 따라 다르게 구성될 수 있다는 것은 이미 언급한 바 있다. 이는 행복과 불행이 영역별로 서로 다른 의미를 가질 수 있으며, 이에 따라 심층적인 분석이 필요하다는 것을 의미한다. 따라서 본 연구는 행복과 불행의 영역별 의미를 살펴보기 위해 K-LIWC 심리 단어 사전의 범주를 활용하여, 행복 및 불행에 대한 감성사전을 구축하고 각각의 K-LIWC 범주를 추출하였다. 감성사전을 통해 행복과 불행에 대한 영역별 K-LIWC 범주를 확인할 수 있으며, 이를 통해 행복 및 불행의 심리적 의미를 파악할 수 있을 것으로 기대하였다.

본 연구에서는 K-LIWC 심리 범주 44개 중 종속적인 10개 범주를 제외한 34개 범주를 활용하였으며, 년도별로 가장 지배적인 속성을 갖는 순서로 K-LIWC 범주를 나열하였다. 또한 각 범주에 따른 상위 5%(+2SD 이상)의 단어 목록이 제시되었다. 또한 각 단어가 문서 내에서 가지는 중요도를 가중하기 위해 TF-IDF 값을 곱하여 누계하였다.

감성 분석은 속성별 단어 수에 따른 가중치를 고려하기 위해 등장 빈도에 log 값을 취하여 계산하였다.

표 5. 행복 관련 기사의 K-LIWC 주요범주 속성 값

개념	범주 수	2011	2012	2013	2014	2015	2016	2017
긍정적인 정서	긍정적인 느낌	0.45	0.47	0.27	0.46	0.43	0.47	0.44
	낙천성/활동성	0.16	0.17	0.12	0.22	0.18	0.20	0.17
부정적인 정서	불안	0.10	0.08	0.04	0.06	0.06	0.06	0.06
	화	0.09	0.08	0.05	0.08	0.07	0.08	0.09
	슬픔/우울	0.09	0.08	0.04	0.08	0.06	0.06	0.06
인지적인 과정	원인	0.22	0.24	0.16	0.22	0.17	0.18	0.17
	사고	0.40	0.40	0.27	0.46	0.43	0.41	0.40
	기대	0.03	0.05	0.03	0.04	0.03	0.04	0.04
	제한	0.13	0.13	0.13	0.14	0.11	0.12	0.11
	추측	0.12	0.12	0.07	0.10	0.10	0.10	0.10
	확신	0.36	0.44	0.30	0.44	0.40	0.43	0.38
감각지각적인 과정		0.39	0.40	0.23	0.41	0.40	0.41	0.38
사회적 과정	체면	0.06	0.07	0.03	0.06	0.05	0.05	0.07
	의사소통	0.78	0.97	0.64	0.92	0.82	0.92	0.87
	타인참조	0.04	0.06	0.03	0.05	0.03	0.04	0.03
	또래친구	0.13	0.16	0.08	0.14	0.14	0.12	0.12
	가족	0.46	0.41	0.24	0.43	0.41	0.39	0.38
	인간	0.35	0.34	0.22	0.37	0.35	0.39	0.37
자기활동	학교	0.85	0.95	0.53	1.00	1.00	1.00	1.00
	직장일	0.76	0.82	0.54	0.79	0.76	0.79	0.80
	성취	0.51	0.65	0.38	0.64	0.54	0.63	0.55
여가활동	집	0.39	0.37	0.36	0.43	0.42	0.50	0.41
	운동	0.30	0.35	0.14	0.37	0.21	0.27	0.21
	TV영화	0.54	0.49	0.22	0.38	0.33	0.38	0.35
	음악	0.42	0.41	0.19	0.51	0.34	0.36	0.32
돈/재정적 이슈		1.00	1.00	1.00	0.94	0.97	0.80	0.76
신체적 상태와 기능	몸상태와 증상	0.47	0.45	0.23	0.38	0.36	0.42	0.37
	성성징	0.10	0.08	0.05	0.06	0.07	0.05	0.05
	식사/음주/다이어트	0.26	0.29	0.13	0.23	0.27	0.22	0.24
	수면/꿈	0.04	0.06	0.03	0.04	0.04	0.04	0.04
형이상학 적 이슈	죽음	0.05	0.05	0.03	0.06	0.04	0.05	0.04
	종교	0.25	0.24	0.12	0.23	0.16	0.18	0.17
속어		0.05	0.04	0.02	0.02	0.02	0.02	0.02
맹세어절		0	0	0	0	0	0	0
상위5% 값		0.80	0.96	0.58	0.93	0.87	0.84	0.82

2011-2017년까지 행복과 관련된 범주들은 [표 5]에 제시되어 있다. 상위 5% 값에 해당되는 범주들은 행복과 관련성이 높은 범주로, 음영으로 표시하였다. 지난 7년 간 행복과 가장 관련성이 높은 범주는 ‘돈/재정적이슈’, ‘학교’, ‘의사소통’으로 나타났다. 즉, 한국인에게서 행복이라는 개념은 ‘돈/재정적이슈’, ‘학교’ 같은 사회경제적 요인들과 밀접한 관련이 있다고 해석해 볼 수 있다. 또한 ‘사회적 과정’ 범주에 속하는 하위범주들 중 ‘의사소통’ 범주는 높은 속성 값을 나타냈지만, 나머지 ‘사회적 과정’ 범주의 하위범주는 전반적으로 점수가 낮았다.

표 6. 불행 관련 기사의 K-LIWC 주요범주 속성 값

개념	범주 수	2011	2012	2013	2014	2015	2016	2017
긍정적인 정서	긍정적인 느낌	0.27	0.28	0.28	0.18	0.36	0.43	0.35
	낙천성/활동성	0.24	0.20	0.13	0.11	0.24	0.21	0.29
부정적인 정서	불안	0.17	0.19	0.18	0.13	0.24	0.26	0.26
	화	0.35	0.33	0.35	0.27	0.52	0.41	0.48
	슬픔/우울	0.26	0.22	0.25	0.20	0.35	0.34	0.25
인지적인 과정	원인	0.20	0.27	0.19	0.19	0.35	0.21	0.32
	사고	0.61	0.36	0.52	0.31	0.46	0.50	0.65
	기대	0.08	0.08	0.07	0.05	0.11	0.10	0.10
	제한	0.29	0.39	0.34	0.22	0.30	0.39	0.51
	추측	0.12	0.16	0.17	0.16	0.17	0.12	0.17
	확신	0.43	0.33	0.47	0.29	0.49	0.48	0.60
감각지각적인 과정		0.46	0.43	0.39	0.45	0.72	0.44	0.58
사회적 과정	체면	0.08	0.17	0.07	0.05	0.11	0.09	0.17
	의사소통	0.60	0.67	0.67	0.40	0.76	1.00	1.00
	타인참조	0.03	0.09	0.10	0.02	0.08	0.08	0.07
	또래친구	0.07	0.16	0.11	0.06	0.18	0.10	0.12
	가족	0.36	0.40	0.30	0.35	0.54	0.50	0.31
	인간	0.29	0.29	0.22	0.27	0.40	0.25	0.29
자기활동	학교	1.00	0.73	0.56	1.00	0.94	0.60	0.53
	직장일	0.70	0.56	0.62	0.45	0.72	0.59	0.72
	성취	0.52	0.57	0.31	0.24	0.48	0.50	0.41
여가활동	집	0.28	0.18	0.22	0.15	0.20	0.18	0.12
	운동	0.08	0.20	0.12	0.21	0.28	0.24	0.29
	TV영화	0.20	0.33	0.31	0.31	0.48	0.42	0.36
	음악	0.34	0.33	0.13	0.13	0.14	0.31	0.39
돈/재정적 이슈		0.72	1.00	1.00	0.58	1.00	0.56	0.64
신체적 상태와 기능	몸상태와 증상	0.57	0.61	0.43	0.46	0.92	0.84	0.81
	성성징	0.12	0.10	0.09	0.10	0.34	0.17	0.12
	식사/음주/다이어트	0.21	0.16	0.14	0.10	0.34	0.15	0.25
	수면/꿈	0.06	0.04	0.06	0.03	0.06	0.03	0.08
형이상학 적 이슈	죽음	0.17	0.09	0.11	0.16	0.22	0.14	0.08
	종교	0.50	0.44	0.23	0.24	0.34	0.30	0.29
속어		0.01	0.08	0.01	0.01	0.02	0.12	0.03
맹세어절		0	0	0	0	0	0	0
상위5% 값		0.71	0.69	0.64	0.50	0.93	0.68	0.75

2011-2017년까지 불행과 관련된 범주들은 [표 6]에 제시되어 있다. 상위 5% 값에 해당되는 범주들은 불행과 관련성이 높은 범주로, 음영으로 표시하였다. ‘돈/재정적이슈’의 경우 5년 동안 연속적으로 불행과 관련된 높은 속성 값을 나타낸 동시에, 행복과도 가장 관련성이 높은 범주 중 하나였다. 마찬가지로, ‘학교’, ‘의사소통’은 행복과 가장 관련이 높은 범주인 동시에 불행과도 가장 관련이 높은 범주였다. 그밖에 ‘몸 상태와 증상’이 불행과 높은 관련성이 있는 범주로 나타났다.

V. 논의 및 결론

5.1 연구 결과의 요약 및 논의

본 연구의 목적은 텍스트 마이닝 분석을 통해 한국 사회에서 행복과 불행이 갖는 심층적인 의미를 탐색하고, 한국인의 행복과 불행의 변화를 파악하는 기초자료를 제공하는 것이다. 이를 위해 2011-2017년 언론사 뉴스 기사를 대상으로 ‘행복’과 ‘불행’에 대한 키워드 분석을 실시하였다. 먼저 ‘행복’과 ‘불행’에 대한 빈도분석을 실시한 후, TF-IDF와 Word2Vec 분석을 통해 데이터의 중요 단어와 키워드 유사 단어를 추출하였다. 마지막으로 K-LIWC 심리단어 사전의 34개 범주와 ‘행복’ 및 ‘불행’ 개념의 관련성을 알아보았다.

첫 번째로, 수집된 키워드 언급 기사 수를 바탕으로 행복과 불행 뉴스에 대한 빈도 분석이 이루어졌다. 그 결과 행복 및 불행 기사들의 빈도는 2011년부터 2017년까지 매우 빠르게 증가했다. 특히 2012년과 2013년은 전년도에 비해 각각 144%와 90% 증가한 것으로 나타났다. 이는 행복과 불행이 최근 들어서 매우 중요한 키워드로 부각되었음을 의미한다. 이러한 시대적 흐름과 맞닿아서, 해외에서도 긍정심리학의 출현과 더불어 다양한 긍정심리학 연구가 실시되고 있고[51], 국내에서도 행복증진 개입활동을 통해 사람들의 행복감을 높이기 위한 연구들이 활발하게 진행되고 있다[52-55]. 앞으로도 행복과 관련된 다양한 연구들이 국내에서 계속해서 실시되어, 사람들의 행복감을 증진시키는 데 기여할 필요가 있다. 또한 2011년 이후 기사 빈도의 급격한 변화는 OECD에서 발표하는 더 나은 삶의 지수(Better Life Index, BLI)의 영향도 있을 수 있다. BLI 지수는 OECD의 정책 추구 방향성을 파악할 수 있는 지표 중 하나이며[56], 제작 이후 지표의 검증 및 개선사항[57][58]과 같은 다양한 논의를 촉발하였다. 따라서 국제적 기준인 BLI가 받아들여지면서 삶의 질에 비교적 쉽게 접근할 수 있게 되었으며, 언론의 언급도 자연스럽게 증가했다고 이해할 수 있을 것이다.

둘째로, 행복 및 불행 관련 기사의 연도별 TF-IDF 주요 단어를 살펴보았다. 이 분석을 통해 2011년도부터 2017년도까지 각각의 년도에서 행복과 불행에 관련된

그 해 특징적인 단어들을 추출할 수 있었다. 행복의 경우 사회적 요인과 관련된 단어가 7년 간 연속해서 나타나고 있었다. 또한, 행복보다 불행에 있어 해당 년도에 사회적으로 큰 이슈가 되었던 사건들에 더 많은 영향을 받은 것으로 나타났다. 또 다르게 주목할 만한 변화는, 최근 들어 SNS가 행복에 영향을 미치는 변인이 되었다는 것이다. ‘인스타그램’은 2015년에서 2017년까지 3년 연속으로 행복과 중요성이 높은 단어로 선정되었다. 이러한 점은 SNS가 행복에 영향을 미치는 주제 중 하나라는 연구 결과들과 일치하는 결과이다[50][59]. 특히, 금희조[59]의 연구에서는 다양한 SNS 유형들에 대한 연구가 수행되었는데, 연구 결과 마이크로 블로그인 트위터의 경우 신뢰 형성을 통해, 페이스북과 같은 네트워킹 사이트는 친교 활동을 촉진시키면서 개인의 행복감을 증진시킨다는 것을 발견하였다. 또 다른 연구에서는, SNS를 통해 사용자들이 자기 개방을 함으로써 삶의 만족에 긍정적인 영향을 받음이 확인되었다[60]. 이는 인터넷 사용이 개인의 사회 활동 참여, 저하, 외로움, 우울감 등을 경험하게 한다는 주장을 하는 기존의 연구들[61-63]과는 상반되는 결과이다. 본 연구의 결과는 SNS 가입자 및 이용자 수가 늘어나는 것과 더불어 SNS 상의 활동이 개인의 행복에 영향을 미칠 수 있음을 시사하며, SNS 활동과 행복감 등의 긍정적 정서와 관련된 연구가 필요함을 시사한다. 차후에 SNS 활동 중 어떠한 요소들이 행복과 관련되는지에 대한 연구가 시행된다면 보다 정확한 결과를 알 수 있을 것으로 보인다.

세 번째로, Word2Vec 분석 결과 2011년부터 2017년도까지 ‘성공’, ‘희망’과 같은 단어들은 지속적으로 행복과 높은 유사성을, ‘불안’, ‘위협’과 같은 단어들은 지속적으로 불행과 높은 유사성을 보였다. 즉, 행복과 관련된 단어는 긍정 정서와, 불행과 관련된 단어는 부정 정서와 좀 더 관련되는 것으로 나타났다. 이러한 결과는 행복이 긍정 정서를 많이 경험하면서 부정 정서를 적게 경험하는 것과 관련된다는 선행 연구들[1][4][64][65]을 지지한다.

마지막으로, LIWC를 통한 심리적 의미 분석 결과 지난 7년 간 행복과 가장 관련성이 높은 범주는 ‘돈/재정

적이슈', '학교', '의사소통'과 같은 사회적, 경제적 요인에 해당하는 범주였다. 이 결과는 주관적 삶의 질이 '돈/재정적이슈', '학교'와 같은 심리적 범주와 관련된다는 선행 연구[39]와 일치한다. 불행의 경우, 년도마다 관련 범주가 상이하였으나 행복과 유사하게 사회적 또는 경제적 요인이 높은 관련성이 있었다. 범주들 중, 특히 '돈/재정적 이슈'의 경우 5년 동안 연속적으로 행복과 불행 모두와 높은 관련이 있다고 보고되었다.

'돈/재정적이슈', '학교', '직장일', '성취'와 같은 외적/물질적 조건에 속하는 범주들이 행복과 불행에 동시에 영향을 많이 미치는 범주였다는 사실은 선행연구에 비추어 볼 때 중요한 의미를 갖는다. Diener 등[66]은 한국인들이 미국, 일본, 짐바브웨 사람들보다 더 물질적 풍요를 중시하며, 물질과 같은 외적 단서에 주의를 기울이기 때문에 더 높은 지위를 추구하고 더 많은 사회 비교를 하게 되어 낮은 행복 수준을 느끼게 된다고 보고하였다. 국내의 선행연구들에서도 '돈'과 같은 경제적 측면이 행복의 중요한 조건임이 나타났다[26][28][33][67]. 또한 Suh[30]는 한국과 같은 집단주의 문화권에서는 소속 욕구가 강하고 사회적 승인이 강조되기 때문에 내적, 주관적 경험보다 외적, 사회적 준거(학벌, 돈 등)에 의해 자신을 평가하는 경향이 강하며, 이것이 행복을 낮추는 중요한 요인으로 작용할 수 있다고 하였다. 유나영[68]의 연구에서도 미국인의 행복 개념에서 나타나지 않는 사회경제적 요인들이 한국인의 행복개념에는 중요한 것으로 나타난 반면, 심리적 요인과 사회적 요인은 미국의 행복 개념에서보다 한국인의 행복 개념에서 축소되어 나타났다.

또한, 국내의 선행연구들과는 달리 가족이나 또래친구 등의 범주는 행복이나 불행에 큰 관련이 없다고 나타난 점도 주목할 만하다. 박선영[9]의 연구에서는 한국 성인이 가장 행복한 때로 가정의 화목을 꼽았고 여가생활 범주에서 '가족과 함께할 때' 가장 행복함을 느낀다고 나타났다. 구계선 등[69]의 연구에서도 가족이나 친구와 사회적 관계 경험을 하는 것이 행복을 유발하는 경험임이 나타났다. 가족, 친구 등의 사회적지지 역시 행복과 관련된다는 결과가 상당히 많으나[7][16][70-72], 본 연구에서는 이러한 결과가 지지되지 않았다. 그 이

유로 본 연구의 자료를 수집한 '뉴스'의 특징을 들 수 있다. 뉴스의 경우, 개인적인 사건이나 일상생활에 관련된 내용보다는 사회적인 이야기를 더 많이 다루게 된다. 이러한 경향으로 인해 가족, 친구 등의 단어가 다루어지는 빈도 자체가 적을 수 있을 것이다.

또 다른 흥미로운 점은 행복과 달리 '몸 상태와 증상'이 2년여 간 불행과 관련이 높은 범주로 나타났다. 이는 주관적 삶의 질이 높은 집단이 '신체적 상태와 기능' 범주 단어의 사용 빈도가 낮다는 결과를 지지하며[39], 그 중에서도 '몸상태와 증상' 범주가 불행과 관련된다는 점을 암시한다. 이는 건강한 상태가 박탈당했을 때, 즉, 몸 상태에 이상이 생기거나 특정 증상이 나타났다를 때는 불행을 느낄 수 있다는 결과를 의미한다.

5.2 연구의 시사점, 한계점과 향후 연구 방향

본 연구의 시사점은 다음과 같다. 우선, 본 연구에서는 기존 K-LIWC의 심리 사전을 확장하여 활용하였다. K-LIWC 사전은 44개의 범주를 가진 16,347개의 심리학적 단어를 포함하는 사전이었다. 그러나 현재 K-LIWC 사전에 포함된 단어들은 신조어와 같은 최근 단어의 사용 경향과 의미를 분석하기에 어려움이 있었다. 따라서 본 연구에서는 이러한 제약점을 딥러닝 학습을 통하여 보완하였고 행복 및 불행에 대한 감성사전을 새롭게 구축했다. 새롭게 만들어진 이 감성사전을 활용하여 차후에 행복 및 불행의 의미에 대한 다양한 분석이 가능할 수 있을 것으로 기대된다.

두 번째로, 본 연구는 한국인의 행복 및 불행에 관한 다년간의 데이터를 살펴보았다는 점에서 의의가 있다. 기존에는 대체로 단기간 수집한 데이터를 활용한 연구들이 주를 이루었다. 그러나 본 연구에서는 시공간적 제약을 받지 않는 빅데이터를 활용하여 2011년도부터 2017년도에 이르는 행복 및 불행 관련 뉴스 기사 총 347,606 건을 분석에 사용하였다. 이를 통해 7년 간 한국 사회에서 행복 및 불행이 갖는 의미를 탐색적으로 살펴볼 수 있었다.

반면, 본 연구는 분석 자료로 '뉴스'라는 매체를 활용하였다는 점에서 필연적인 제한을 갖는다. 우선, 뉴스는 특정 사회나 문화권의 현상들을 다룬다는 점에서 객관

적인 데이터로 유용하나, 이로 인해 정서 관련 단어의 언급 빈도는 적은 편이다. 또한, 뉴스는 보다 거시적 차원에서의 사회적 현상을 다루기 때문에 개인이 매일 체험하고 살아가는 일상생활과의 밀접성은 다소 떨어지는 편이기도 할 것이다. 그렇기 때문에 본 연구에서는 긍정적 정서나 부정적 정서 등과 같은 정서 변인들과 행복 및 불행의 관련이 낮은 편으로 나타났다. 그러나 개인적인 삶의 반영 비율이 보다 높은 SNS 연구의 경우 정서적 단어 언급과 개인의 주관적 삶 만족이나 불행과 관련이 높게 나타났다는 선행 연구 결과가 존재하였다[36][38]. 따라서 추후에 시행되는 연구에서는 이러한 점을 고려하여 개인 블로그나 기타 SNS를 활용한 연구를 해 본다면 또 다른 결과를 발견할 수 있을 것이다. 단순히 행복 및 불행과 긍정 또는 부정 정서와의 관련성을 살펴 볼 수도 있겠으며 정서표현의 차이 혹은 빈도가 행복 및 불행과 어떤 관계를 보이는지 등과 관련해서도 의미 있는 결과를 얻을 수 있을 것으로 기대된다. 뿐만 아니라, 행복이 부정적 정서의 부재와도 관련이 되는지, 또는 불행이 긍정적 정서의 부재와 관련이 되는지 등 다양한 흥미로운 결과를 도출할 수 있을 것이다.

다음으로, K-LIWC를 활용한 행복 및 불행과 관련된 기존의 선행연구들이 부족하여 탐색적 연구에 그쳤다는 점 역시 본 연구의 제한점이다. 본 연구에서는 K-LIWC의 34개 범주들을 모두 활용한 포괄적인 접근 방법을 사용하였고 행복 및 불행과 관련된 기술적 수준에 그쳤다. 이로 인해 어떤 특성의 변인들이 행복 또는 불행에 영향을 미치는지, 또는 영향을 받는지와 관련해서는 결과를 도출하지는 못하였다. 이후 연구들에서는 본 연구를 바탕으로 하여 행복 및 불행과 관련된 K-LIWC 범주를 특정적으로 활용할 수 있을 것이다. 추후 연구들에서 행복 또는 불행과 연관이 있는 범주들이 추려지면 각 년도 마다 행복 또는 불행과 관련된 범주들로 행복 및 불행의 프로파일을 매년 도출할 수도 있을 것이다. 또한, 시계열 분석이나 패널 분석 등의 연구 방법을 활용하여 행복 또는 불행의 종단적 의미 변화를 살펴본다면 한국 사회에서의 행복과 불행에 대한 보다 풍부한 결과를 얻을 수 있을 것으로 기대된다.

* K-LIWC 한국어 심리 단어 사전에 대한 이해와 연구 결과의 의미를 해석하는데 도움을 주신 아주대학교 심리학과 김영준 박사님에게 감사드립니다.

참 고 문 헌

- [1] E. Diener, "Subjective well-being," *Psychological Bulletin*, Vol.95, pp.542-575, 1984.
- [2] E. Diener, E. M. Suh, R. E. Lucas, and H. L. Smith, "Subjective well-being: Three decades of progress," *Psychological bulletin*, Vol.125, No.2, pp.276-302, 1999.
- [3] M. A. Busseri, "Examining the structure of subjective well-being through meta-analysis of the associations among positive affect, negative affect, and life satisfaction," *Personality and Individual Differences*, Vol.122, pp.68-71, 2018.
- [4] C. D. Ryff, "Happiness is everything, or is it? Explorations on the meaning of psychological well-being," *Journal of Personality and Social Psychology*, Vol.57, No.6, pp.1069-1081, 1989.
- [5] C. L. M. Keyes, "Social well-being," *Social Psychology Quarterly*, Vol.61, No.2, pp.121-140, 1998.
- [6] C. L. M. Keyes and J. L. Magyar-Moe, "The measurement and utility of adult subjective well-being," In S. J. Lopez & C. R. Snyder (Eds.), *Positive psychological assessment: A handbook of models and measures*, (pp.411-425), Washington, DC, US: American Psychological Association, 2003.
- [7] M. Seligman, *Positive psychology, positive prevention, and positive therapy*, In C. Snyder & S. Lopez (Eds). *Handbook of positive psychology*, New York: Oxford, 2002.
- [8] M. Seligman, *긍정심리학 [Authentic Happiness]* (김인자 역), 안양: 물푸레 (원서 2002 출판), 2011.

- [9] 박선영, “한국 성인이 지각한 행복의 요인,” 한국심리학회 학술대회 자료집, 제2017권, 제8호, pp.184-184, 2017.
- [10] 박영신, 김의철, 한기혜, 박선영, “청소년이 지각한 행복의 수준과 요인,” 청소년학연구, 제19권, 제9호, pp.149-188, 2012.
- [11] 전기숙, “청소년의 행복 요인 탐색,” 청소년학연구, 제23권, 제10호, pp.411-429, 2016.
- [12] 정혜숙, 김영희, “아동이 지각한 행복 결정요인,” 청소년학연구, 제21권, 제7호, pp.89-118, 2014.
- [13] 서경현, 이경순, “대학생의 생활 스트레스와 사회비교 및 주관적 웰빙 간의 관계,” 한국심리학회지: 건강, 제15호, 제2권, pp.327-343, 2010.
- [14] 조공민, 이덕남, “청소년의 스트레스가 주관적 안녕감에 미치는 영향,” 지역사회연구, 제22권, 제4호, pp.57-74, 2014.
- [15] 구분용, 유제민, “성격 및 환경요인과 행복과의 관계,” 한국심리학회지: 건강, 제6권, 제1호, pp.1-9, 2005.
- [16] 명수정, 강승호, “중학생의 행복에 대한 사회적 지지와 학업 성취 및 낙관성의 효과,” 교육평가연구, 제24권, 제3호, pp.619-643, 2011.
- [17] 조유진, 한지숙, “저소득가정 아동의 사회적 지지가 학교 적응에 미치는 영향: 행복감의 매개효과를 중심으로,” 아동과 권리, 제14권, 제4호, pp.541-562, 2010.
- [18] 이종한, 박은아, “내-외적 자기개념, 행복조건, 사회비교와 자기존중감의 관계,” 한국심리학회지: 문화 및 사회문제, 제16권, 제4호, pp.423-445, 2010.
- [19] 전혜진, 유미숙, “청소년의 행복에 영향을 미치는 자아존중감, 진로성숙도 및 미시체계 변인의 경로분석,” 청소년학연구, 제22권, 제6호, pp.75-104, 2015.
- [20] 김정실, 박영신, “성격과 행복의 관계에 대한 연구,” 한국심리학회 학술대회 자료집, 제2016권, 제8호, pp.312-312, 2016.
- [21] 인보영, 김인옥, “아동의 성격강점이 학교생활 행복에 미치는 영향,” 한국웰니스학회지, 제7권, 제1호, pp.81-95, 2012.
- [22] C. L. M. Keyes, “Gender and subjective well-being in the United States: From subjective well-being to complete mental health,” *Psychology of stress*, pp.1-15, 2005.
- [23] J. Helliwell, R. Layard, and J. Sachs, *World Happiness Report 2017*, New York: Sustainable Development Solutions Network, 2017.
- [24] 구재선, 서은국, “한국인, 누가 언제 행복한가?,” 한국심리학회지: 사회 및 성격, 제25권, 제2호, pp.143-166, 2011.
- [25] S. Lyubomirsky, K. M. Sheldon, and D. Schkade, “Pursuing happiness: The architecture of sustainable change,” *Review of General Psychology*, Vol.9, No.2, pp.111-131, 2005.
- [26] 이제현(책임프로듀서), 행복해지는 법 2편 행복의 비밀코드, [TV 다큐멘터리], 서울: KBS TV, 2010.
- [27] 구재선, 서은국, “왜 한국 대학생이 미국 대학생보다 불행한가? 상대적 외적 가치, 사회적 지원, 사회비교의 영향,” 한국심리학회지: 사회 및 성격, 제29권, 제4호, pp.63-83, 2015.
- [28] 박정현, 서은국, “사람의 내-외적인 모습에 두는 상대적 비중과 행복관과의 관계,” 한국심리학회지: 사회 및 성격, 제19권, 제4호, pp.19-31, 2005.
- [29] 박영신, 김의철, “한국 사회에서 삶의 질을 구성하는 요인들에 대한 탐구,” 한국심리학회지: 문화 및 사회문제, 제12권, 제5호, pp.1-28, 2006.
- [30] E. M. Suh, “Downsides of an overly context-sensitive self: Implications from the culture and subjective well-being research,” *Journal of Personality*, Vol.75, pp.1321-1343, 2007.
- [31] E. M. Suh and J. Koo, Comparing subjective well-being across cultures and nations: The “what” and “why” questions., In M. Eid & R. J. Larsen (Eds.), *The science of subjective well-being*, NY: Guilford Press, pp.414-427,

- 2008.
- [32] 박영신, 김의철, “심리적, 관계적, 경제적 자원 - 한국인의 행복에 어떠한 영향을 미치는가?,” 한국심리학회지 : 문화 및 사회문제, 제15권, 제1호, pp.95-132, 2009.
- [33] 한준, 김석호, 하상응, 신인철, “사회적 관계의 양면성과 삶의 만족,” 한국사회학, 제48권, 제5호, pp.1-24, 2014.
- [34] M. Luhmann, “Using Big Data to study subjective well-being,” *Current Opinion in Behavioral Sciences*, Vol.18, pp.28-33, 2017
- [35] C. Chung and J. W. Pennebaker, “The psychological functions of function words,” *Social communication*, Vol.1, pp.343-359, 2007.
- [36] P. Liu, W. Tov, M. Kosinski, D. J. Stillwell, and L. Qiu, “Do Facebook status updates reflect subjective well-being?,” *Cyberpsychology, Behavior, and Social Networking*, Vol.18, No.7, pp.373-379, 2015.
- [37] S. Collins, Y. Sun, M. Kosinski, D. Stillwell, and N. Markuzon, “Are you satisfied with life?: Predicting satisfaction with life from facebook,” In *International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction*, Springer, Cham, pp.24-33, Mar. 2015.
- [38] M. Settanni and D. Marengo, “Sharing feelings online studying emotional well-being via automated text analysis of Facebook posts,” *Frontiers in psychology*, Vol.6, p.1045, 2015.
- [39] 김경일, 배진희, 김영진, 김동근, “주관적 삶의 질과 언어 사용의 관계성 분석,” 한국산학기술학회 논문지, 제12권, 제11호, pp.4875-4880, 2011.
- [40] 우병현, “네이버의 뉴스스탠드 도입으로 보는 언론사의 향후 과제,” *언론중재*, 제33권, 제2호, pp.60-75, 2013.
- [41] 이창환, 심정미, 윤애선, “언어적 특성을 이용한 ‘심리학적 한국어 글분석 프로그램 (KLIWC)’ 개발 과정에 대한 고찰,” *인지과학*, 제16권, 제2호, pp.93-121, 2005.
- [42] 김경일, 임종섭, 이창환, “심리학 언어분석 프로그램을 이용한 인문학 연구 탐색,” *인문논총*, 제37호(단일호), pp.131-146, 2015.
- [43] G. Salton and C. S. Yang, “On the specification of term values in automatic indexing,” *Journal of documentation*, Vol.29, No.4, pp.351-372, 1973.
- [44] W. Zhang, T. Yoshida, and X. Tang, “A comparative study of TF*IDF, LSI and multi-words for text classification,” *Expert Systems with Applications*, Vol.38, No.3, pp.2758-2765, 2011.
- [45] T. Mikolov, K. Chen, G. Corrado, and J. Dean, “Efficient estimation of word representations in vector space,” *arXiv preprint arXiv:1301.3781*, 2013.
- [46] D. Alessia, F. Ferri, P. Grifoni, and T. Guzzo, “Approaches, tools and applications for sentiment analysis implementation,” *International Journal of Computer Applications*, Vol.125, No.3, 2015.
- [47] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, “Learning representations by back-propagating errors,” *nature*, Vol.323, No.6088, p.533, 1986.
- [48] 송해엽, 양재훈, “포털 뉴스 서비스와 뉴스 유통 변화,” *한국언론학보*, 제61권, 제4호, pp.74-109, 2017.
- [49] 김우주, 김동희, 장희원, “Word2vec 을 활용한 문서의 의미 확장 검색방법,” *한국콘텐츠학회논문지*, 제16권, 제10호, pp.687-692, 2016.
- [50] 이연경, 송의령, 이소현, 김희웅, “텍스트마이닝 기법을 이용한 국가간 행복과 불행 토픽 비교 분석: 한국, 미국, 영국, 브라질,” *한국경영정보학회 학술대회*, pp.738-749, 2017.
- [51] M. E. P. Seligman, *Authentic Happiness: Using the new positive psychology to realize your potential for Lasting fulfillment*, New York: Free Press(김인자 역, 긍정심리학, 서울.

- 물푸레, 2006), 2004.
- [52] 노지혜, 이민규, “나는 왜 감사해야 하는가?: 스트레스 상황에서 감사가 안녕감에 미치는 영향,” 한국심리학회지: 임상, Vol.30, No.1, pp.159-183, 2011.
- [53] 이명자, 신현숙, 염시창, “대학생의 자원과 심리적 안녕감의 관계: 목표추구의 매개효과,” 청소년학연구, Vol.16, No.9, pp.53-79, 2009.
- [54] 황진석, 이명자, “자기설정목표표기, 자기설정목표, 목표전념, 학업성취 및 주관적 안녕의 관계,” 교육심리연구, Vol.24, No.2, pp.397-421, 2010.
- [55] 소연희, “학습자 동기 특성 변인들이 고등학생들의 주관적 안녕감에 미치는 영향 - 성취목표지향성, 자기효능감, 자기결정성, 학습동기 및 학업성취를 중심으로-,” 교육심리연구, Vol.21, No.4, pp.1007-1028, 2007.
- [56] 광숙영, “OECD 의 사회정책관련 변화와 한국 사회정책의 과제,” 한국사회정책, Vol.18, No.4, pp.299-332, 2011.
- [57] 이내찬, “OECD 국가의 삶의 질의 구조에 관한 연구,” 보건사회연구, 제32권, 제2, pp.5-40, 2012.
- [58] 차용진, “삶의 질 측정에 관한 실증적 연구,” 한국공공관리학보, 제27권, 제2호, pp.1-26, 2013.
- [59] 금희조, “소셜 미디어 시대, 우리는 행복한가? : 소셜 미디어 이용이 사회 자본과 정서적 웰빙에 미치는 영향,” 한국방송학보, 제25권, 제5호, pp.7-48, 2011.
- [60] 이규동, 이재은, “자아 개념 명확성과 자기 개방이 소셜 네트워크 사이트에 대한 애착과 삶의 만족에 미치는 영향 : 싸이월드 미니홈피를 중심으로,” 대한경영학회, 제27권, 제11호, pp.1845-1861, 2014.
- [61] J. E. Katz, R. E. Rice, and P. Aspden, “The Internet, 1995-2000: Access, civic involvement, and social interaction,” American behavioral scientist, Vol.45, No.3, pp.405-419, 2001.
- [62] R. Kraut, T. Mukhopadhyay, J. Szczypula, S. Kiesler, and W. Scherlis, “Communication and information: Alternative uses of the Internet in households,” In Proceedings of the CHI 98 (pp. 368-383), New York: ACM, 1998.
- [63] N. H. Nie and L. Erbring. “Internet and society: a preliminary report,” IT&Society, Vol.1, No.1, pp.275-283, 2002.
- [64] D. Nettle, *Happiness*, Oxford: Oxford University Press, 2005.
- [65] 이보라, 박재황, “대학생의 긍정정서, 몰입, 삶의 의미와 행복 간의 구조적 관계,” 현실치료연구, 제6권, 제1호, pp.1-16, 2017.
- [66] E. Diener, E. M. Suh, C. Kim-Prieto, R. Biswas-Diener, and L. S. Tay, “Unhappiness in South Korea: Why it is high and what might be done about it,” 한국심리학회 연차학술대회 자료집, pp.1-23, 2010.
- [67] 박영신, 김의철, 박선영, “성인이 지각한 행복의 수준과 요인 : 가정, 직장, 여가생활을 중심으로,” 인간발달연구, 제20권, 제2호, pp.153-184, 2013.
- [68] 유나영, 정여진, 김비아, 정영숙, 신현정, “한국인의 행복 개념 탐색연구,” 한국민족문화, 제55권, pp.197-230, 2015.
- [69] 구재선, 김의철, “한국인의 행복 경험에 대한 토착문화심리학적 접근,” 한국심리학회지: 문화 및 사회문제, 제12권, 제2호, pp.77-100, 2006.
- [70] S. Cohen and T. A. Wills, “Stress, social support, and the buffering hypothesis,” Psychological bulletin, Vol.98, No.2, pp.310-357, 1985.
- [71] 김숙연, 조규관, “중학생의 학업적 실패내성, 심리적 성장환경, 사회적 지지가 학교행복감에 미치는 영향,” 학습자중심교과교육연구, 제16권, 제2호, pp.259-279, 2016.
- [72] 김정엽, “여성노인의 사회적 지지와 삶의 만족도와의 관계,” 복지행정논총, 제20권, 제2호, pp.29-62, 2010.

