

빅데이터 분석기법을 활용한 아파트 가격 관련 뉴스 기사의 극성 분석

조상연¹, 홍은표^{2,3*}

¹상명대학교 일반대학원 부동산학과 박사과정,
²상명대학교 경제금융학부 교수, ³상명부동산연구소 소장

A Study on the Polarity of Apartment Price News Using Big Data Analysis Method

Sang-Yeon Cho¹, Eun-Pyo Hong^{2,3*}

¹Ph.D. Candidate, Department of Real Estate, The Graduate School, Sangmyung University,

²Professor, Department of Economics and Finance, Sangmyung University,

³Director, Sangmyung Institute for Real Estate Study

요약 본 연구는 빅데이터 분석 방법인 오피니언 마이닝을 사용하여 아파트 가격 관련 뉴스 기사의 극성을 확인하는 연구로 자료는 2012년, 2018년 2년간 네이버에 게시된 인터넷 뉴스 기사를 사용하였다. 감성분석 모형을 모델링하고 주제 지향형 감성사전 구축 방법을 제안하였다. 제안한 감성분석 모형을 통해 분석한 결과, 아파트 가격이 상승하는 시기에는 사회적 이슈 선정에 있어서 언론사의 성향에 따라 차이가 있는 것을 확인하였고 정부와 동일한 성향의 언론사에서 긍정 기사가 많은 것을 확인하였다. 부동산 분야에서 사용할 수 있는 감성분석 모형을 제시하고 부동산 관련 비정형 데이터의 극성을 분석하였다는 것에 의의가 있다. 향후 다양한 분야에 접목하기 위해서는 주제별 감성사전을 구축해야 하며 다양한 비정형 데이터를 수집하고 수집 기간을 확장하는 것이 필요하다.

주제어 : 부동산, 아파트, 빅데이터, 오피니언 마이닝, 감성분석 모형

Abstract This study confirms the polarity of news articles on apartment prices using Opinion Mining which has widely been used for a big data analysis. The analyses were carried out utilizing internet news articles posted on the Naver for two years: 2012 and 2018. We proposed a sentiment analysis model and modeled a topic-oriented sentiment dictionary construction methods. As a result of analyzing the proposed sentiment analysis model, it was confirmed that there was a difference according to the tendency of the media companies in selecting social issues at the time of rising apartment prices. At the same time, we were able to find more affirmative articles in the media companies which share similar sentiment with the government in charge. In this paper, we proposed a sentiment analysis model that can be used in real estate field and analyzed the polarity of unformatted data related to real estate. In order to integrate them into various fields in the future, it is necessary to build the sentiment dictionaries by themes, as well as to collect various unformatted data over extended periods.

Key Words : Real Estate, Apartment Big Data, Opinion Mining, Sentiment Analysis Model

*Corresponding Author : Eun-Pyo Hong(ephong@smu.ac.kr)

Received July 10, 2019

Accepted September 20, 2019

Revised August 24, 2019

Published September 28, 2019

1. 서론

1.1 연구의 배경 및 필요성

인터넷, 모바일, 통신장비 등 통신 기술의 발전에 힘입어 정보 기술(Information Technology)은 눈부신 발전을 이루었다. 정보 기술의 발전은 의사결정의 패러다임을 데이터 주도형 의사결정으로 바꾸고 있으며 데이터를 어떻게 사용하느냐에 따라 개인, 기업 더 나아가서는 국가에 엄청난 기회와 막대한 부를 가져올 수 있다는 사실은 다양한 사건을 통해서 사람들에게 부각되었다. 알리바바의 CEO 마윈은 “20년 동안 지속된 정보 기술 시대가 저물고 향후 30년간 데이터 기술 시대가 열린다”라고 언급하였다[1].

데이터 중에서도 4차 산업혁명과 함께 주목받고 있는 빅데이터는 세계 각국에서 경쟁 역량 차별화를 위한 필수 자산으로 인식되고 있다[2].

빅데이터는 의료, 미용, 패션, 물류, 금융, 공공 등 다양한 산업으로 확산되고 있으며 전통산업인 부동산 산업에서도 해외를 중심으로 변화가 활발히 진행 중이다. 2006년 설립한 미국 1위 부동산 플랫폼 회사인 ‘질로우 그룹(Zillow Group)’은 빅데이터와 머신러닝으로 주택 가격을 산출하는 ‘제스트메이트(Zestmate)’라는 차별화된 포인트로 2017년 8월 기준으로 5년 평균 매출 증가율이 69.1%이며 기업가치는 6.6억 달러로 평가받고 있다[3].

우리나라에서는 온라인 부동산 중개 앱 등이 확산되는 추세이나 해외 기업들처럼 빅데이터를 활용한 전문적인 부동산 분석 기술은 도입되지 않고 있다.

빅데이터의 형태는 데이터의 정형화 정도에 따라 정형 데이터, 반 정형 데이터, 비정형 데이터로 분류할 수 있으며 이 중 모바일, SNS(Social Network Service), IoT(Internet of Things) 등의 비정형 데이터가 급증하고 있다. 비정형 데이터는 정해진 구조가 없는 텍스트 문서, 이미지, 동영상 등과 같이 기계적으로 파악하는 것이 불가능한 데이터를 말한다[4].

주식시장 관련 최근 연구들은 비정형 데이터인 뉴스 기사에 대한 분석을 통해 주가를 예측하는 연구로 집중되고 있다. 기업에 대한 긍정, 부정 뉴스 기사가 주가의 상승, 하락에 영향을 미칠 것이라고 가정하고 뉴스 기사로부터 어휘 패턴을 학습하여 60%대의 높은 예측 정확도를 보였다[5].

국내 주식시장에서는 언론사인 NEWSIS와 구 한국증권 전산회사인 Koscom가 협력하여 온라인상의 기업

에 대한 긍정, 부정 여론을 분석하여 1(가장 낮음)~ 7(가장 높음) 단계로 등급화한 빅데이터 시장심리 지수인 MSI(Market Sentiment Index)를 2017년 8월 1일부터 투자자들에게 제공하고 있다.

부동산 분야에서도 시장 참여자의 의사결정에 효율성을 높이거나 다양한 부동산 문제를 해결하기 위해서 비정형 데이터를 포함한 대규모 데이터를 분석할 수 있는 빅데이터 연구의 필요성이 요구되고 있으나 부동산 관련 빅데이터에 관한 연구는 미비하다. 특히 부동산 관련 비정형 데이터에 대한 연구는 다른 분야에 비해서 많이 부족한 상황이며 비정형 데이터의 수집 방법, 분석 기법 등에 대한 다양한 연구가 필요하다.

비정형 데이터 중 매스미디어의 경제 뉴스는 시장 참여자의 심리와 경제에 상당한 영향을 끼친다[6]. 언론사의 성향에 따라서 시장 참여자의 심리에 영향을 끼치는 것은 다를 수 있다. 이는 동일한 부동산에 대한 뉴스 기사라도 언론사에 따라 기사의 내용은 긍정적이거나 부정적일 수 있기 때문이다. 언론사의 성향에 따라 어젠다 설정이나 뉴스 프레임의 영향력은 다르며 여론 형성이나 여론 갈등에 결정적인 변수로 작용하고 있다[7].

비정형 데이터 분석에 장점이 있는 빅데이터 분석 기법을 활용하여 부동산 관련 뉴스 기사를 분석한다면 언론사별 성향에 따라 뉴스 기사의 극성을 수치화할 수 있으며 정량적으로 분석할 수 있을 것으로 예상된다.

1.2 연구의 목적 및 내용

본 연구의 목적은 빅데이터 분석 방법인 오피니언 마이닝을 활용하여 언론사별 부동산 시장 관련 뉴스 기사의 극성을 확인하는 것이다.

연구의 내용을 구분하면 첫째, 언론사별 기사 내용을 수집하고 분석하기 위한 감성분석 모형을 모델링하고 시스템을 구축한다. 둘째, 구축된 시스템을 사용하여 언론사별 기사 내용을 수집한다. 셋째, 수집된 데이터를 기반으로 주제 지향 감성사전을 구축하고 언론사별로 기사의 극성을 태깅한다. 마지막으로, 극성 값을 분석하여 정량적으로 비교한다.

연구의 방법은 다음과 같다. 2장에서는 선행연구를 통해 오피니언 마이닝의 활용 및 연구 방향에 대해 고찰한다. 3장은 비정형 데이터 수집, 콘텐츠 분류, 형태소 분석, 극성 태깅 및 스코어링의 기능을 수행하는 감성분석 모형을 모델링하고 감성분석 모형에서 사용하는 주제 지향 감성사전 구축 방법을 제안한다. 4장은 실제로 구축된

시스템에서 수집된 데이터를 실증 분석한다. 마지막으로 5장에서는 연구 결과를 요약하고 시사점을 제시한다.

2. 선행연구 고찰

2.1 오피니언 마이닝에 관한 연구

빅데이터 분석 방법 중 하나인 오피니언 마이닝은 감성분석(Sentiment Classification)이라고도 하며 텍스트의 주관적인 정보나 극성, 극성의 정도를 분석하기 위해 이용된다[8,9]. 감성분석이 기업들에게 주목받기 시작한 것은 1995년부터 2010년 출생한 세대를 일컫는 ‘플로팅 세대’들이 소비의 중심으로 바뀌기 시작하면서부터이다. 플로팅 세대는 하나의 콘텐츠에 길게 집중하지 못하고 여러 가지 정보를 동시다발적으로 수용하는 세대로 콘텐츠뿐만 아니라 직장과 거주지도 유목민처럼 여기저기 옮겨 다니는 세대를 말한다. 이 세대들은 합리적인 소비보다는 감성을 자극하는 콘텐츠를 구매의 최우선 순위로 여긴다[1]. 이러한 시대적 변화에 기업들은 앞 다투어 소비자의 감성을 분석하고자 연구하고 있다.

기업 이외에 오피니언 마이닝을 활발히 연구하고 있는 분야는 주식시장이다. 뉴스와 주가의 밀접한 관계에 주목하여 오피니언 마이닝 기반의 투자 의사결정 모형, 주식시장에 특화된 주제 지향 감성사전 구축 방안, 오피니언 마이닝과 기계학습을 이용한 주가예측 시스템 등 다양한 방향으로 연구가 진행되고 있다[10-12].

오피니언 마이닝과 기계학습을 이용한 주가예측 시스템의 연구 결과는 과거에 많이 사용하고 있는 수학적, 통계적 기법의 예측 방법보다 정확함을 확인하였다[12].

정치 분야에서도 오피니언 마이닝 연구는 활발히 진행되고 있다. 미국 대통령 선거 때 트럼프 캠프를 지원한 케임브리지 애널리티카(CA) 스캔들은 유권자의 감성을 분석하여 변화시킨 예로 거론된다. 데이터 기반 정치 컨설팅 기업인 CA의 요청으로 페이스북에 CA에서 개발한 심리테스트 서비스를 이용할 수 있는 게임 애플리케이션을 올렸고 가입자 27만 명뿐만 아니라 가입자의 친구들까지 포함한 5,000만 명의 개인 정보를 수집한 것이다. CA는 수집된 유권자의 개인 정보(페이스북 댓글, 공유한 게시물, 좋아요 클릭, 위치정보 등)를 이용하여 트럼프의 선거운동을 도왔고 부정적인 인식이 있는 유권자들을 대상으로 ‘친 트럼프 마케팅’을 진행하여 선거 결과에 영향을 주었을 것이라는 의혹을 받았다[13].

국내의 정치분야도 오피니언 마이닝 연구는 진행되고 있다. SNS를 통해 대중들이 의사소통하고 정보를 공유한다는 것에 착안하여 SNS 사용자의 의견들을 텍스트 마이닝과 오피니언 마이닝 방법을 적용하여 분석하였고 지방선거와 19대 대선 등의 선거 결과와 비교하여 상당히 일치함을 확인하였고 여론조사 결과보다 정확함을 확인하였다[14,15].

국내 부동산 분야에서 오피니언 마이닝을 사용한 연구는 텍스트 마이닝과 오피니언 마이닝을 사용한 부동산 분야 빅데이터 감성 분석 모형을 제안한 연구가 있다[6].

본 연구에서는 부동산시장에서 사용할 수 있는 감성분석 모형을 모델링하고 부동산시장에 특화된 감성사전을 주제 지향 방안으로 구축하고자 한다. 제안된 빅데이터 감성분석 모형과 주제 지향 감성사전은 시장 참여자들의 감성을 분석하는 연구에 도움이 될 것으로 예상된다.

3. 감성분석 모형

3.1 감성분석 모형 구조

본 연구에서 제안하고자 하는 감성분석 모형은 경정의 외(2016) 연구의 부동산 분야 빅데이터 감성분석 모형을 수정 보완한 것으로, 감성사전 구축 방안과 기사의 극성 값 산출 방식에 차이가 있다[6]. 감성분석 모형의 구조는 Fig. 1과 같으며 5단계로 동작한다.

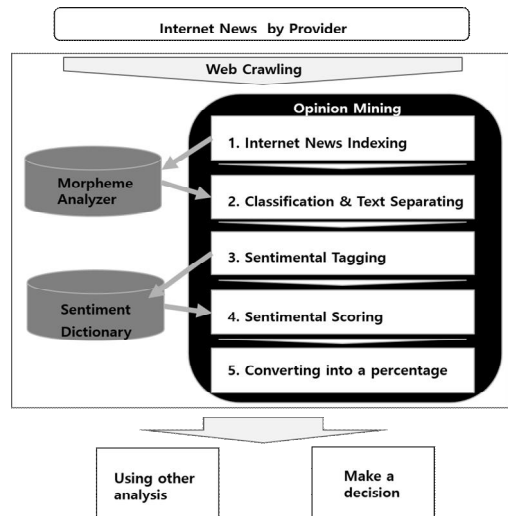


Fig. 1. Sentiment Analysis Model

첫 번째로 포털 사이트에서 특정 검색어로 인터넷 뉴스를 검색한다. 검색된 인터넷 뉴스 기사의 URL(Uniform Resource Locator)을 저장한다.

두 번째로 수집된 URL에서 사용하지 않는 부분은 제외하고 뉴스 기사 관련 정보와 기사 내용을 분리하여 저장한다. 저장된 기사 내용의 형태소를 분석하기 위해서 형태소 분석기를 사용하였다. 형태소 분석기는 한글의 체언, 용언, 부사, 관형사 등의 다양한 품사들을 구분할 수 있다[10]. 본 연구에서는 기사 내용의 극성을 판단하기 위해서 명사와 동사를 추출하여 연구에 활용하였다.

세 번째로 개발된 감성사전을 기반으로 각 명사와 동사에 극성 값을 붙인다.

네 번째로 극성 값이 붙은 문장에 가중치 모델을 적용하여 극성의 정도를 추가하였다.

마지막으로 아래 수식을 적용하여 기사별 극성 값을 100분비로 변환한다.

$$X = \frac{Count_p - Count_N}{Count_T} \times 100 \quad (\text{식1})$$

X는 기사의 극성 값을 100분비로 변환한 값이다. $Count_p$ 는 기사 내 긍정 명사와 동사의 수, $Count_N$ 는 기사 내 부정 명사와 동사의 수, $Count_T$ 는 기사 내 긍정과 부정을 의미하는 명사와 동사의 총 수를 나타낸다.

X가 긍정인 경우는 양(+)의 값을 부정인 경우는 음(-)의 값을 가지며 긍정 또는 부정의 백분비 값을 제공한다.

3.2 주제 지향 감성사전 구축

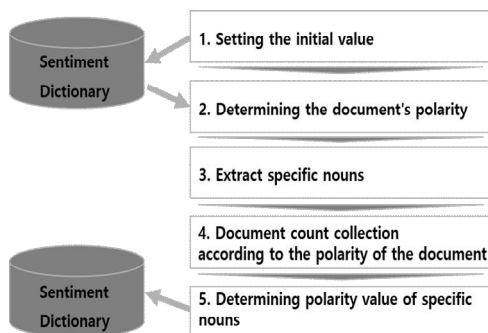


Fig. 2. Sentiment Dictionary Production Procedure

본 연구에서 제안하는 주제 지향 감성사전 구축방안은 유은지 외(2013) 연구의 주식 시장에 특화된 감성사전 구

축 방안을 부동산 시장에 맞게 수정 보완한 것으로 특정 명사 추출방식과 극성 값 계산식에서 차이가 있다[11].

오피니언 마이닝에서 핵심요소인 감성사전은 어휘별로 긍정, 부정의 극성을 규정해 놓은 사전을 의미한다[11].

부동산 분야에 특화된 주제 지향 감성사전 구축은 과정은 Fig. 2와 같으며 5단계로 진행된다.

첫 번째로 범용 감성사전을 사용하여 본 연구의 주제 지향 감성사전 초깃값을 설정한다. 긍정을 나타내는 단어는 1로 부정을 나타내는 단어는 -1로 설정한다. 두 번째로 첫 번째 작업을 진행한 감성사전을 기반으로 특정 주제로 검색된 기사의 긍정과 부정을 판단하여 저장한다. 세 번째로 저장된 기사에서 범용 감성사전에 없는 명사들을 주제 관련 키워드와 최소 빈도수 등의 기준으로 필터링하여 추출한다. 네 번째로 해당 명사가 긍정 기사에 많이 등장하였는지 부정 기사에 많이 등장하였는지 집계한다. 마지막으로 아래 수식을 통해 특정 명사의 극성 값을 정한다[16].

t = 명사

$Document_p(t)$ = 명사 t 를 포함한 긍정문서수

$Document_n(t)$ = 명사 t 를 포함한 부정문서수

$$Sentiment\ Score(t) = \frac{[Document_p(t) - Document_n(t)]}{MAX[Document_p(t), Document_n(t)]} \quad (\text{식2})$$

지속적으로 주제 지향 감성사전의 데이터가 축적되어 주제별 특정 명사들의 극성 값이 보완된다면 주제 지향 감성사전의 정확도는 높아질 것이며 다양한 주제로 분석이 가능할 것이다.

4. 실증 분석

본 연구에서는 부동산 시장 내 공동주택인 아파트의 가격 상승과 가격 하락 두 가지의 검색어로 포털사이트인 네이버에 게시된 인터넷 뉴스 기사를 수집하였다. 공간적 범위는 전국으로 선정하였고 시간적 범위는 2012년 1월 1일부터 2012년 12월 31일까지, 2018년 1월 1일부터 2018년 12월 31일까지 총 2년간으로 정하였다. 시간적 범위를 2012년과 2018년으로 한정한 이유는 한국감정원에서 제공하는 아파트 실거래 가격 지수를 전국 기준으로 2008년부터 2018까지 10년간 확인한 결과 2012년에는 아파트 실거래 가격 지수가 하락하였고 2018년에는 아파트 실거래 가격 지수가 상승하였기 때

문이다[17]. 다른 이유는 집권한 정부의 성격이 다르기 때문이다. 2012년은 아파트 가격 하락으로 기사를 수집 하였고 2018년은 아파트 가격 상승으로 기사를 수집하였다. 언론사는 보수적 성향의 동아일보와 진보적 성향의 한겨레신문을 선정하였다[7].

2012년도 아파트 가격 하락으로 수집된 기사의 건수는 동아일보 70건, 한겨레 신문 31건이었고 2018년도 아파트 가격 상승으로 수집된 기사의 건수는 동아일보 145건, 한겨레신문 69건이었다.

부동산 시장 관련 인터넷 기사에 자주 사용되는 단어들은 정부 정책, 주택 가격 등과 영향 관계가 있다[18]. 본 연구에서는 부동산 시장 관련 인터넷 기사에서 자주 사용되는 단어들이 정부 정책과 주택가격에 영향이 있는 것에 착안하여 언론사별로 자주 사용되는 단어들에 차이점이 있는지 확인하기 위해서 빈도수를 집계하였다.

Table 1은 2012년도 아파트 가격 하락으로 수집된 언론사별 기사들에서 언급된 단어들의 빈도수를 높은 순서대로 1위부터 20위까지 작성한 내용이다.

Table 1. Frequency of Content by media in 2012

Rank	Donga		Hankyoreh	
	Content	Frequency	Content	Frequency
1	apartment	423	price	155
2	housing	363	apartment	132
3	price	262	decrease	114
4	property	252	market	103
5	decrease	223	housing	100
6	market	211	seoul	98
7	seoul	203	property	90
8	reconstruction	123	capital area	85
9	capital area	121	region	69
10	loan	116	deal	64
11	sales	111	home lease	63
12	household	97	loan	58
13	region	92	reconstruction	55
14	deal	91	gangnam	51
15	finance	89	demand	43
16	assets	87	housing price	43
17	complex	84	sales	42
18	gangnam	75	household	39
19	gyeonggi	74	increase	39
20	home lease	71	gyeonggi	37

Table 2는 2018년 아파트 가격 상승으로 수집된 언론사별 기사들에서 언급된 단어들의 빈도수를 높은 순서로 작성한 내용이다.

Table 2. Frequency of Content by media in 2018

Rank	Donga		Hankyoreh	
	Content	Frequency	Content	Frequency
1	apartment	932	housing	614
2	seoul	719	price	578
3	price	716	apartment	421
4	housing	650	seoul	344
5	increase	545	property	305
6	property	518	increase	291
7	market	437	market	283
8	region	437	public notice	271
9	complex	423	government	257
10	household	401	housing price	218
11	housing sales	378	region	216
12	reconstruction	362	reconstruction	173
13	development	315	gangnam	168
14	government	284	homeowner	144
15	gangnam	281	hold	137
16	housing price	280	measure	131
17	detached	256	city	118
18	supply	201	burden	111
19	schedule	196	development	110
20	business	188	house	108

2012년도 아파트 가격 하락 기사는 동아일보와 한겨레신문 모두 가격, 하락, 전세, 재건축, 지역, 강남 등 아파트 가격 하락과 재건축에 관한 단어들이 비슷하게 언급되었다.

2018년도 아파트 가격 상승 기사에서는 언론사별로 언급되는 단어들에서 차이를 보였다. 동아일보는 분양, 공급, 예정 등 신규로 공급되는 아파트에 관한 단어가 언급되었고 한겨레신문에서는 공시, 보유, 대책, 부담 등 정부 정책과 관련된 단어가 언급되었다.

위 결과로 아파트 가격이 하락하는 시기에 작성된 기사는 동일한 사회적 이슈를 다루고 있지만 아파트 가격이 상승하는 시기에는 사회적 이슈 선정에 있어서 언론사의 성향에 따라 차이가 있는 것을 확인할 수 있다.

3장에서 제안한 감성분석 모형을 사용하여 언론사별 기사의 극성 값을 분석하기에 앞서 가격 인상과 관련된 단어와 동사는 부정의 값으로 처리하였고 가격 인화와 관련된 명사와 동사는 긍정의 값으로 처리하였다. 이처럼 인상은 부정의 값으로 인하는 긍정의 값으로 설정한 이유는 문재인 정부의 부동산 정책 목표는 서민 주거안정과 주거복지 향상을 최우선 가치로 삼고, 다주택자나 투기적 수요를 차단하는 것이기 때문이다. 실제로 문재인

정부는 2017년 8·2대책과 2018년 9·13대책 등 고강도 대책을 통해 투기수요를 억제하고 주택시장의 가격 안정을 추진하였다.

분석 결과를 보면 2012년도 동아일보의 총 기사 수 70건 중 부정 기사는 23건이었고 긍정 기사는 47건이었다. 한겨레신문의 총 기사 수 30건 중 부정 기사는 17건, 긍정 기사는 13건이었다. Table 3.은 2012년도 언론사별 긍정 부정 기사의 비율을 나타낸다.

2018년도 동아일보의 총 기사 수 146건 중 부정 기사는 136건이었고 긍정 기사는 10건이었다. 한겨레신문의 총 기사 수 60건 중 부정 기사는 55건이었고 긍정 기사는 5건이었다. Table 4는 2018년도 언론사별 긍정과 부정 기사의 비율을 나타낸다.

Table 3. Positive and Negative Articles ratio by media in 2012

	Donga	Hankyoreh
Positive	67.14%	43.33%
Negative	32.86%	56.67%

Table 4. Positive and Negative Articles ratio by media in 2018

	Donga	Hankyoreh
Positive	6.85%	8.33%
Negative	93.15%	91.67%

2012년은 보수 성향의 이명박 정부가 집권한 시기였고 Table 3을 보면 보수 성향의 동아일보가 진보 성향의 한겨레신문보다 긍정 기사의 비율이 높은 것을 확인할 수 있다. 2018년은 진보 성향의 문재인 정부가 집권하고 있는 시기이기 때문에 진보 성향의 한겨레신문이 보수 성향의 동아일보보다 긍정 기사의 비율이 높다.

위 결과로 정부와 동일한 성향의 언론사에서 긍정 기사가 많은 것을 확인할 수 있었다.

감성사전을 구축하는 과정에서 긍정과 부정의 뜻을 가진 단어들을 수집하였는데 2012년도에는 한겨레신문의 총 기사 수가 동아일보보다 적음에도 부정적인 뜻을 가진 단어의 총 수는 동아일보 104개, 한겨레신문 114개로 한겨레신문이 동아일보보다 10개가 더 많았다. 정부와 반대되는 성향의 언론사 기사에서 부정적인 뜻을 가진 단어들의 수가 많았으며 자극적인 단어들의 사용도 많았다.

5. 결론

빅데이터 분석방법인 오피니언 마이닝 기법을 활용하여 언론사별 부동산시장 관련 뉴스 기사의 극성을 본 연구에서 확인하였다.

모델링한 주제 지향형 감성분석 모형을 시스템으로 개발하였다. 시스템은 웹 데이터수집, 콘텐츠 분류, 형태소 분석, 극성 태깅 및 스코어링의 기능을 수행하는 4가지 모듈로 구성된다. 웹 데이터수집 모듈은 JavaScript와 Java 프로그래밍 언어를 사용하여 개발하였고, 콘텐츠 분류와 극성 태깅 모듈들은 Java를 사용하여 개발하였다. 형태소 분석기는 오픈소스로 제공되는 Komoran 3.0버전을 사용하여 개발하였다.

DataBase는 Mysql을 사용하여 구축하였고 프로그램 개발 툴은 이클립스를 사용하였다.

분석을 위해 2년간(2012년, 2018년) 네이버에 게시된 인터넷 뉴스 기사를 자료로 사용하였다. 동아일보와 한겨레신문에서 아파트 가격 상승, 하락을 주제로 수집된 뉴스 기사를 각각 사용하였다.

언론사별로 사회적 이슈를 선정함에 있어서 차이점을 확인하기 위해 기사에서 사용하는 명사의 빈도수를 집계하였고 2012년도 아파트 가격 하락 시기에는 동아일보와 한겨레신문 모두 아파트 가격 하락과 재건축에 관한 단어들 이 비슷하게 언급되었다. 2018년도 아파트 가격 상승 시기에는 언론사별로 언급되는 단어에서 차이를 보였는데 동아일보는 신규로 공급되는 아파트에 관한 단어들 이 언급되었고 한겨레신문에서는 정부 정책과 관련된 단어들 이 언급되었다.

구축된 시스템을 사용하여 언론사별 극성을 분석한 결과를 보면 2012년도 긍정 기사의 비율은 동아일보 67.14%, 한겨레신문 43.33%이었고 ; 부정 기사의 비율은 동아일보 32.86%, 한겨레신문 56.67%였다. 2018년도 긍정 기사의 비율은 동아일보 6.85%, 한겨레신문 8.33%이었고 ; 부정 기사의 비율은 동아일보 93.15%, 한겨레신문 91.67%로 나타났다. 정리하면 동일한 사회적 이슈로 기사를 작성하더라도 정부와 동일한 성향의 언론사에서 긍정 기사가 많은 것을 확인할 수 있었다.

감성사전을 구축하는 과정에서 긍정과 부정의 뜻을 가지는 단어들을 수집하였는데 정부와 반대되는 성향의 언론사 기사에서 부정적인 뜻을 가진 단어들이 더 많은 것을 확인하였다.

본 연구의 결과는 주제 지향 감성사전을 부동산 분야에 접목한 부분에 의의가 있으며 부동산 분야에 특화된

주제 지향 감성사전을 구축하고 모델링한 감성분석 모형을 시스템으로 개발하고 실증한 것에 의미가 있다.

향후 언론사의 범주와 데이터 수집 기간을 확장하여 본 연구 결과의 객관성과 강건성(robustness)을 검증할 필요가 있다. 아울러 다양한 비정형 데이터 분석을 통해 본 연구의 주제인 부동산 시장에 특화된 주제 지향 감성사전의 정확도를 제고할 수 있을 것이다.

빅데이터에 관한 기술은 분석과 처리, 정보, 데이터, 예측, 서버, 서비스, 구축 등 여러 방면에서 연구가 진행 중이며 정부는 빅데이터의 활용을 권장하기 위해서 정부 3.0이라는 사업을 통해 Open API와 공공데이터들을 개방하고 공유하고 있다[19,20].

부동산 분야에서도 새로운 빅데이터에 관한 기술과 정부에서 제공하는 공공데이터를 활용하여 부동산 분야에 특화된 빅데이터에 관한 연구가 필요하다. 빅데이터를 이용하면 분석할 수 없었던 데이터들을 수치화하여 확인할 수 있으며 이를 기반으로 정확하고 신속하게 부동산 관련 이슈들을 처리할 수 있을 것으로 기대된다.

REFERENCES

- [1] N. D. Kim. (2018). *Trend Korea 2019*. Seoul : The Window of Future.
- [2] Korea Development Bank. (2016. Oct). The Rise of Big Data Industry and Implications. *Monthly KDB Survey Report*, 731, 82-102.
- [3] Korea Development Bank. (2017. Sep). Change The Trend of The Real Estate Market with Big Data. *Weekly KDB Report, Issue Brief*, 8-11.
- [4] J. I. Kyung. (2014). Big Data Utilization Scheme of The Real Estate Sector and Policy Recommendations. *Journal of The Korea Real Estate Management*, 10, 65-67.
- [5] W. Paik, M. H. Kyoung, K. S. Min, H. R. Oh, C. Lim & M. S. Shin. (2007). Multi-Stage News Classification System for Predicting Stock Price Changes. *Journal of The Korean Society for Information Management*, 24(2), 123-141.
- [6] J. I. Kyung & K. C. Lee. (2016). Development of Sentiment Analysis of Real Estate Big data by Using Textmining. *Housing Studies Review*, 24(4), 115-136.
- [7] S. H. Yang. (2012). *The Changing News Framing in Korean Journalism -Focused on Comparing The Reports about KORUS FTA on The Chosun Ilbo, The Seoul Shinmun and The Hankyoreh*. Doctoral Dissertation. Sungkunkwan University, Seoul.
- [8] Wilson, Theresa, Janyce Wiebe & Rebecca Hwa. (2004). Just how mad are you? Finding Strong and Weak Opinion. *aaai*, 4.
- [9] Esuli, Andrea & F. Sebastiani. (2005). Determining The Semantic Orientation of Terms through Gloss Classification. *ACM*.
- [10] Y. S. Kim. (2012). *News Big Data Opinion Mining Model for Predicting KOSPI Movement*. Doctoral Dissertation. Kookmin University, Seoul.
- [11] E. J. Yu, Y. S. Kim, N. G. Kim & S. R. Jeong. (2013). Predicting The Direction of The Stock Index by Using a Domain-Specific Sentiment Dictionary. *Korea Intelligent Information Systems Society*, 19(1), 95-110.
- [12] K. B. Kim. (2016). *News Big Data Opinion Mining Model for Predicting KOSPI Movement*. Doctoral Dissertation. Soongsil University, Seoul.
- [13] J. H. LEE. (2018). Cambridge Analytical Scandal Over Facebook. *Economy Chosun*, 245.
- [14] W. S. Kim, J. H. Lee, J. W. Park & J. H. Choi. (2014). A Technique of The Approval Rating Analysis for Political Party Using Opinion Mining. *Korean Institute of Information Technology*, 12(10), 133-141.
- [15] Y. N. Lee, E. J. Choi & M. J. Kim. (2018). Analysis of The Influence of Presidential Candidate's SNS Reputation on Election Result: Focusing on 19th Presidential Election. *Journal of Digital Convergence*, 16(2), 195-201.
- [16] S. W. Kim & N. K. Kim. (2013). A Study on The Effect of Using Sentiment Lexicon, *Korea Intelligent Information Systems Society*, 121-128.
- [17] Korea Appraisal Board. (2019. Feb). *Transaction-Based Price Indices for The Multi-Unit Housing Market*, 116072, 1-2.
- [18] M. H. Jang & H. S. Kim. (2019). A Research on Fluctuations of Housing Prices Using Text Mining. *Journal of The Korea Housing Association*, 30, 35-42.
- [19] J. C. Choi. (2018). Big Data Patent Analysis Using Social Network Analysis. *Journal of the Korea Convergence Society*, 9(2), 251-257.
- [20] Y. B. Cho, S. H. Woo & S. H. Lee. (2013). In Small and Medium Business the Government 3.0-based Big Data Utilization Policy. *Journal of Convergence for Information Technology*, 3(1), 15-22.

조 상 연(Sang-Yeon Cho)

[상화]



- 2006년 2월 : 호서대학교 컴퓨터공학과(공학사)
- 2008년 2월 : 호서대학교 컴퓨터공학과(공학석사)
- 2018년 12월 : 상명대학교 부동산학과 박사과정 수료
- 관심분야 : 주택, 부동산, 빅데이터, 인공지능

공지능

· E-Mail : noa1001@gmail.com

홍 은 표(Eun-Pyo Hong)

[상임]



- 1998년 ~ 2014년 : OECD 비즈니스 마이크로 데이터 Unit 실장
- 2014년 ~ 현재 : 상명대학교 경제금융학부 정교수
- 2019년 ~ 현재 : 상명부동산연구소 소장
- 관심분야 : 부동산, 경제, 금융, 계량통계
- E-Mail : ephong@smu.ac.kr