



방송뉴스 감성지수와 서울시 주택매매가격의 상관 및 인과관계 분석*

이재수**, 박재수***

요약

이 연구의 목적은 부동산 관련 방송뉴스 감성지수가 서울시 아파트 매매가격지수와 상관 및 인과관계가 있는지 밝히는 것이다. 지상파 3사의 부동산 관련 방송뉴스를 웹 크롤링으로 수집·분류·전처리하고, 토픽분석, 텍스트랭크 알고리즘, TF-IDF(Term Frequency-Inverse Document Frequency) 분석, 나이브 베이즈 분류 모델 등 기계학습 기법을 적용하여 월별 감성지수를 산출하였다. 분석 결과, 방송 감성지수는 서울시 아파트 매매가격지수와 양의 상관관계를 나타냈다. 또한 방송 감성지수는 매매가격지수에 1개월부터 5개월 시차까지 유의미한 인과관계를 나타냈다. 이 연구는 텍스트 마이닝과 비정형 빅데이터를 활용하여 부동산시장의 메커니즘을 설명 및 예측할 수 있음을 시사한다. 기계학습 등 인공지능 기법이 부동산 산업에 활용될 수 있는 점도 시사한다.

주제어: 주택매매가격, 방송뉴스, 감성지수, 교차상관관계, 그랜저인과관계

1. 서론

우리나라 가계가 소유한 자산 중 가장 높은 비중을 차지하는 것은 부동산 관련 자산이다. 부동산이 차지하는 비중은 선진국과 비교해도 매우 높은 수준이다. 주택은 사람이 생활하는데 필수적인 요소이지만, 가계자산 중 부동산의 비중이 높은 우리나라에서 주택은 주거 이외에 재산 증식을 위한 중요한 투자재의 의미도 있다.

주택가격 등 주택시장의 흐름을 설명하고 예측

하는 것은 우리나라 대부분의 가구의 관심사이다. 주택 및 부동산시장의 예측은 일반적으로 통계학적 접근 및 계량경제학적 접근을 통해 이루어졌다. 개별 주택가격의 결정 메커니즘과 요인별 중요도의 설명 및 예측은 헤도닉 이론 등에 기반하여 미시적, 통계학적 분석방법을 통해 주로 이루어졌다. 거시적 주택시장의 변화에 대한 설명 및 예측은 주로 패널 또는 시계열 분석방법을 통해 물가, 금리, 경제성장률 등 거시경제변수가 주택의 수요 또는 공급에 미치는 영향을 분석하였다(박재수·이재

* 본 논문은 박재수(2020)의 '주택시장 예측을 위한 부동산 감성지수 개발 연구'를 수정 및 보완하여 재구성한 것임.

** (제1저자) 강원대학교 부동산학과 부교수, E-mail: jslee25@kangwon.ac.kr

*** (교신저자) 한일네트웍스 상무, E-mail: okparkjaesoo@naver.com

수, 2019).

선행연구는 미시적 차원과 거시적 차원에서 외생변수가 주택가격 등 주택시장의 변화를 설명하고, 단기 및 중장기 예측을 위한 변수와의 관계를 설정하는데 기여하였다. 그러나 최근 부동산시장 참여자의 주관적 또는 심리적 변수에 대한 고려가 필요하다는 주장이 제기되고 있다. 국토연구원은 주택, 토지를 포함한 부동산시장 소비심리지수가 2011년부터 조사하여 매월 발표하고 있다(국토연구원, 2020). 그러나 이 조사는 비용이 많이 소요되고 빠르게 변화하는 현대사회에서 즉시성이 떨어진다. 최근 인공지능과 빅데이터 기술이 발전함에 따라 이 기술을 활용하여 부동산 관련 심리지수를 산출하여 이용할 수 있다.

이 연구는 부동산 관련 방송뉴스 스크립트를 이용하여 산출한 방송뉴스 감성지수가 서울시 아파트 매매가격지수와 유의미한 상관 및 인과관계가 있는지 밝히는데 목적이 있다. 2012년 1월부터 2018년 12월까지 지상파 3사(KBS, MBC, SBS)의 부동산 관련 방송뉴스를 웹 크롤링(Web Crawling)을 통해 수집·분류하고 기계학습 기법을 활용한 감성분석을 통해 감성지수를 산출하였다. 서울시 아파트 매매가격지수는 KB 국민은행 자료를 이용하였다.

II. 선행연구 검토

1. 언론과 부동산시장의 관계

선행연구는 뉴스 등 언론과 부동산시장의 관계를 분석하고, 비정형 빅데이터를 활용하여

주택 등 부동산시장을 예측한 연구와 부동산시장에 참여하는 소비자의 심리가 주택 등 부동산시장에 미치는 영향을 분석한 연구로 구분할 수 있다. 언론과 부동산시장의 관계에 대한 연구는 대부분 신문기사와 같은 비정형 빅데이터를 활용하고 있다.

Gayer and Viscusi(2002)는 유해 폐기물 부지에 대한 뉴스기사 보도가 주택가격에 미치는 영향을 분석하였다. 분석 결과, 뉴스기사 진당 약 100~200달러의 주택가격이 증가하였다. 이는 위해에 대한 주민의 인식을 낮추고, 폐기물의 정화에 대한 기대를 증가시키기 때문이라고 주장하였다. Sun et al.(2014)은 온라인 뉴스와 구글 검색엔진의 데이터를 결합하여 부동산 가격예측의 가능성을 분석하였다. 분석 결과, 온라인 뉴스 이용자의 이용행태가 반영된 검색엔진 데이터가 포함된 모형의 부동산 가격의 예측력이 우수한 것으로 나타났다.

국내 연구로 김진유(2006)는 신문 기사를 수집하여 투기가 포함된 언론기사가 부동산 가격의 변화에 미치는 영향을 분석하였다. 분석 결과, 투기와 관련된 기사가 많으면 부동산 가격은 오르는 양의 관계가 나타난다고 주장하였다. 우윤석·이은정(2011)은 언론보도의 빈도와 주택가격의 변화 관계를 분석하였다. 서울 강남의 아파트와 관련된 기사의 빈도수가 증가하면 일정 시차를 두고 기타 지역의 주택가격에 영향을 주는 것으로 나타났다.

진창하·Paul(2012)은 신문기사의 내용을 기반으로 한 주택시장의 인식과 투자심리가 주택가격에 어떤 영향을 미치는지 탐색하였다. 미국의 애틀랜타 CMSA를 사례로 한 연구에서

긍정적 용어를 사용한 기사에 비해 부정적 용어를 사용한 뉴스기사의 내용이 주택의 가격변화에 더 큰 영향을 미치는 것으로 분석되었다. 김대원·유정석(2016)은 뉴스기사의 어조와 주택가격의 지속기간의 관계를 분석하였다. 분석 결과, 기사의 긍정 또는 부정적 어조가 주택가격의 상승 또는 하락 국면의 지속기간에 유의미한 영향을 미친다고 주장하였다.

2. 심리지수와 주택시장의 관계

부동산시장에 대한 심리가 주택시장에 미치는 영향을 실증한 연구는 다양하다. 우선 최희갑·임병분은 주택가격의 전망이 주택가격과 경기에 미치는 영향을 분석하였다. 시계열분석을 통해 주택시장의 투자자의 부정적 태도를 반영하는 가격전망지수가 시차를 두고, 주택가격과 경기에 유의미한 영향을 미친다고 주장하였다.

김대원·유정석(2013)은 주택가격의 심리적 태도가 매매거래량에 미치는 영향을 분석하였다. 분석 결과, 주택가격에 대한 심리적 태도 변수는 일정한 시차를 두고 주택매매를 위한 거래량에 유의미한 영향을 주는 것으로 나타났다. 김리영·안지아(2013)도 이와 유사하게 주택시장 소비자의 주택가치에 대한 전망이 주택가격 및 거래에 미치는 영향을 탐색하였다. 주택시장에서 주택가치에 대한 향후 전망이 주택가격보다는 주택의 거래량에 유의미한 영향을 미친다고 주장하였다. 주택가치에 대한 긍정적 전망이 거래량의 증가에 영향을 미침으로써 주택가격의 증가에는 큰 영향을 미치지 않는 것

으로 판단된다.

조태진(2014)은 부동산시장의 심리지수가 주택시장에 미치는 영향을 분석하였다. 서울 등 우리나라 7개 대도시의 거시경제변수와 심리지수가 주택시장에 미치는 영향을 패널모형을 구축하여 분석하였다. 분석 결과, 부동산전망지수는 단기와 중기 등 모든 시차에서 주택시장의 변화에 유의미한 영향을 미치며 반응하는 것으로 나타났다.

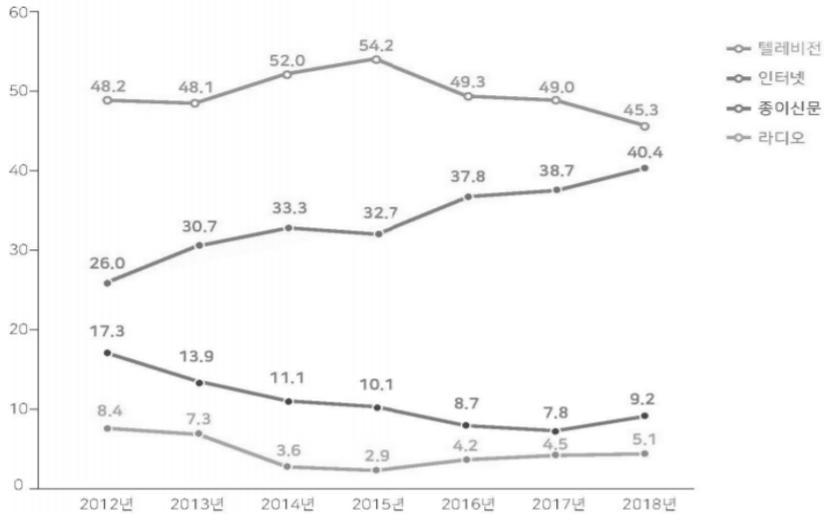
유한수·정재호(2015)는 주택시장에서 주택의 매매가격지수와 소비자심리지수 간의 관계를 실증하였다. 소비심리지수와 주택 매매가격지수가 양방향의 유의미한 그랜저인과관계를 나타냈다. 소비심리지수는 주택 매매가격지수의 증가 또는 감소에 원인이 되고, 반대로 주택 매매가격지수의 변화는 소비심리지수의 변화에 유의미한 영향을 미치는 것으로 해석된다.

노민지·유선종(2016)은 소비자 심리와 아파트 매매가격 간의 관계를 분석하였다. 인터넷을 통해 검색한 아파트 매매가격에 대한 검색량과 국토연구원이 발표하는 주택매매가격 소비심리지수의 관계를 분석한 결과, 이들은 상호간에 시차없이 영향을 미치는 동시성의 관계가 나타났다.

Ⅲ. 연구 방법

1. 분석 자료

최근 조사한 연도별 매체부문의 영향력을 살펴보면, <그림 1>과 같이 TV가 미치는 영향력



자료: 여론집중도조사위원회. (2018).

〈그림 1〉 연도별 매체부문의 영향력 및 가중치

이 가장 높다. 그러나 TV의 영향력은 2015년 54.2%를 정점으로 서서히 감소하는 반면, 인터넷 매체의 영향력은 조사된 2012년 이후 지속적으로 급격히 증가하고 있음을 알 수 있다. 이에 따라 온라인 뉴스와 방송 사이의 가중치의 차이가 크게 줄어들어서 뉴스 이용자는 인터넷과 TV에 대한 의존도와 영향력을 비슷한 수준에서 평가하고 있는 것으로 나타났다. 그러나 여전히 TV를 통한 뉴스의 전달은 사회적 파급력이 높다.

인터넷 뉴스는 스마트폰의 급속한 보급으로 TV 다음으로 많이 뉴스를 접하는 매체가 되었다. 이로 인해 우리나라 사람들이 뉴스 등 미디어를 소비하는 형태의 변화를 가져왔다. 한국언론진흥재단이 실시한 ‘2018년 언론수용자 의식조사’에 따르면, 뉴스 소비에 가장 많이 이용하는 매체는 TV, 스마트폰, 컴퓨터 순으로

나타났다.

2015년 뉴스를 보기 위해 TV를 시청하는 응답 비율은 40.5%이고, 스마트폰을 이용하는 비율은 34.9%로 나타났다(〈그림 2〉 참고).

2018년 기준, 뉴스를 보기 위해 TV를 시청하는 응답 비율은 40.5%이고, 스마트폰을 이용하는 비율은 34.9%로 나타났다. 스마트폰의 보급 증가로 인해 뉴스의 소비도 인터넷과 스마트폰을 이용하는 비중이 점점 더 증가하지만, TV를 이용한 뉴스의 소비와 TV 방송뉴스의 사회적 파급력은 여전히 매우 높음을 알 수 있다.

이 연구에서 부동산 방송뉴스는 지상파 3사(KBS, MBC, SBS)에서 부동산이 포함된 뉴스를 추출하였다. 해당 사이트의 방송뉴스 기사 스크립트를 웹 크롤링(Web Crawling)한 후, ‘아파트’와 ‘매매’가 포함된 뉴스 기사를 2차로 분류하였다. 방송뉴스의 추출 건수는 〈표 1〉과 같다.

〈표 1〉 방송뉴스 기사 및 문장 추출 건수

방송사	기사 건수	문장 개수
KBS	3,276	42,460
MBC	2,001	20,033
SBS	2,886	42,460

2. 분석 방법

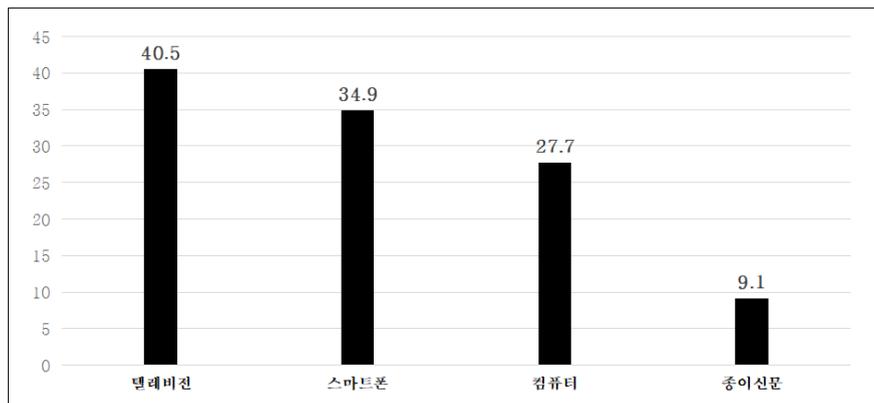
방송뉴스 기사를 이용하여 부동산 감성지수를 산출하는 절차는 뉴스기사 전처리, 기계학습을 활용한 감성분석, 각 문장에 대한 감성지수 산출, 최종 월별 감성지수 산출로 구성된다. 우선 방송뉴스 기사를 수집하기 위해 방송 3사의 웹사이트에서 부동산 관련 텍스트 기사 중에서 아파트와 매매가 모두 포함된 문서를 웹 크롤링한다. 수집한 뉴스기사에서 불용어와 특수 문자 제거 등과 같은 전처리 절차를 시행한다.

토픽분석을 이용하여 수집된 방송 뉴스 빅데이터의 내용을 8개 토픽, 토픽당 30개 단어를

추출한다. 토픽분석에서 추출한 단어가 포함된 문장 9,600개 무작위 추출하여 감성사전을 구축하기 위한 표본 데이터를 구성한다. 연구자 외 2인이 상의하여 이 문장에 대한 긍정과 부정을 판별한다. 추출한 문장과 판별된 긍정 및 부정의 문장 자료를 이용하여 텍스트랭크(TextRank) 알고리즘을 통해 문장에 포함된 단어 간의 관계를 분석하고, 이를 활용하여 감성사전을 만든다.

다음 단계는 TF-IDF를 이용하여 분석을 위해 추출한 모든 기사의 문장에 나타난 단어의 점수를 계산한다. 계산된 단어의 점수를 나이브 베이즈 분류 모델의 입력값으로 투입한다. 모델을 이용하여 문장에 긍정과 부정 극성을 부여하고, 해당 문장의 긍정과 부정 가중치를 산출한다.

마지막 단계에서 감성지수 도출 공식에 따라 월별로 수집된 방송뉴스 전체에 대한 방송 감성지수를 산출한다. 산출된 방송 감성지수와 서울시 월별 아파트 매매가격지수 사이의 시계



자료: 한국언론진흥재단 미디어연구센터. (2015).

〈그림 2〉 뉴스 소비에 주로 이용하는 매체

열적 교차상관관계와 그랜저인과관계 분석을 실시하고 이를 해석한다.

1) 토픽 분석과 텍스트랭크 분석

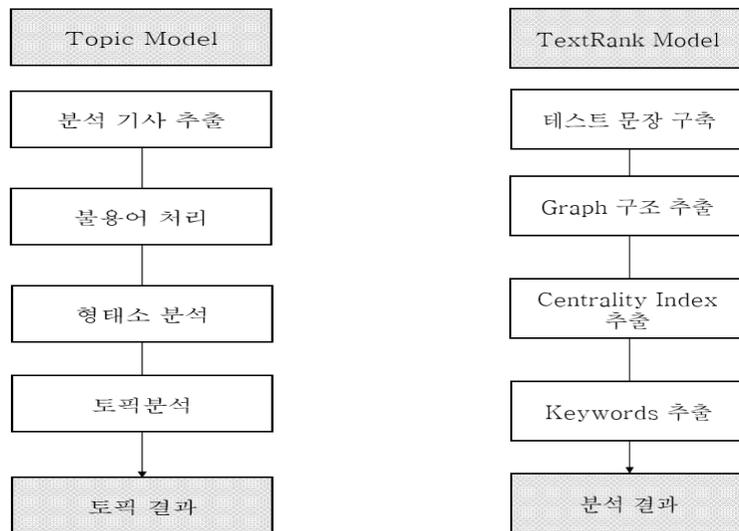
토픽 모델(topic model)은 텍스트 문서 집합의 추상적 주제(topic)를 발견하기 위한 통계적 방법 모델로, 텍스트 본문의 숨겨진 의미구조를 발견하기 위해 사용되는 텍스트 마이닝 기법 중 하나이다(안정욱 외, 2015). 대표적인 방법은 잠재디리클레 할당(latent dirichlet allocation, LDA)이다. LDA는 문헌 내의 숨겨져 있는 주제들을 찾아내는 알고리즘이다. 문헌, 단어 등 관찰 변수를 통해 문헌의 구조와 같은 보이지 않는 변수를 추론한다. 전체 문서 집합의 주제들과 문서별 주제 비율, 각 주제에 포함될 단어들의 분포를 알 수 있다(차윤정 외, 2015; Blei, 2012).

토픽 모델은 확률적 토픽 모델이라고도 불린다.

광범위한 텍스트 본문의 잠재적 의미 구조를 발견하기 위한 통계적 알고리즘을 가리키는 의미이다. 이 연구에서 적용한 토픽 모델은 <그림 3>과 같다.

텍스트랭크(TextRank)는 텍스트를 대상으로 한 그래프 기반의 순위 모델이다. 구글의 검색 엔진의 하나인 페이지랭크(PageRank)에 가중치를 부여한 응용 알고리즘으로 문서 요약 및 단어 추출에 사용된다(배영준 외, 2018). 텍스트랭크 알고리즘은 그래프 기반의 랭킹 알고리즘에서 데이터를 단어로 취급하여 그래프 기반으로 단어 사이의 연관관계를 구성하여 키워드의 중요도를 계산한다.

텍스트랭크 알고리즘은 문서를 그래프로 표현하고, 정점과 정점 간의 간선을 사용하여 요도를 결정하는 알고리즘이기 때문에 그래프를 구성하는 단계가 중요하다. 텍스트랭크는 텍스트 A에서 텍스트 B로 연결된 링크 하나를 텍스트 A가 텍스트 B에게 던지는 한 표로 해석하여 특정 텍스트의



<그림 3> 토픽분석과 텍스트랭크 분석 방법

득표수를 기준으로 중요도를 평가한다. 이 연구에서 적용한 방법은 <그림 4>와 같다.

2) TF-IDF 분석과 나이브 베이즈 분석

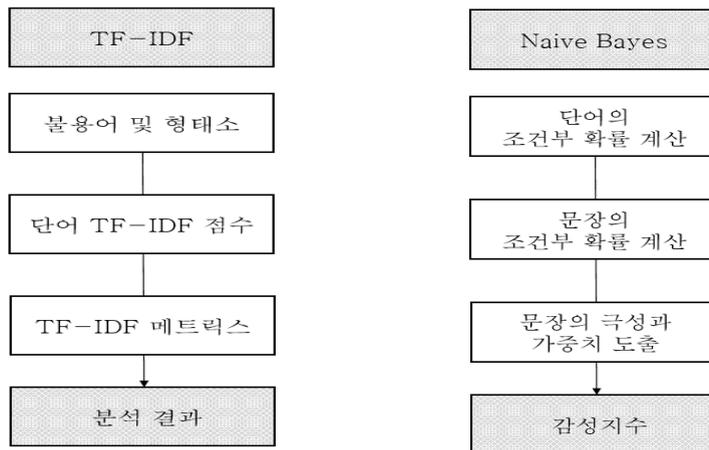
TF-IDF는 텍스트 마이닝에서 중요한 단어를 추출하여 단어사전을 구축할 때 많이 쓰이는 방법으로 여러 문서들에서 출현한 각각의 단어들의 중요성을 점수화하여 준다. TF(term frequency)는 문서들에 나오는 단어의 빈도수를 의미하고, IDF(inverse document frequency)는 역문서 빈도라는 뜻으로서 특정한 문서에만 많이 나오는 정도를 의미한다(Stevens et al., 2014). 일반적으로 이 두 개를 곱한 값을 TF-IDF로 정의하는데, 본 연구에서는 전처리한 텍스트에서 TF-IDF로 다시 한번 필터링하여 중요 단어를 추출하고, 동시에 점수화한 행렬 값을 수치화된 단어(feature)로 사용하여 나이브 베이즈 모델을 만드는데 재료로 사용된다.

IDF는 전체 문서의 수를 해당 단어가 포함된 문서들의 수로 나눈 값에 로그를 취한 값으로 특별한

문서에만 나타나는 단어일수록 점수가 높게 나온다. 따라서 TF와 IDF를 곱한 값은 특정 문서 내에서 빈도수가 높을수록 전체 문서에서는 빈도수가 낮을수록 높아지는 경향이 있어 단어의 중요도를 평가하는 좋은 척도가 된다. TF-IDF 가중치에 대한 연구는 여러 학술분야에서 인용되어 왔으며, 다른 텍스트 마이닝 기법을 통해 좀 더 의미 있는 결과를 도출하는데 쓰일 수 있다(조은경, 2015).

나이브 베이즈 분류는 텍스트 분석, 다중 클래스 분류(multi-classification), 스팸 필터링(spam filtering), 감정분석(sentiment analysis) 등에 사용된다(Lee et al., 2010). 나이브 베이즈(Naïve Bayes) 알고리즘은 속도가 빠른 것이 장점이다. 이 연구에서 적용한 나이브 베이즈 분석은 <그림 4>와 같다.

본 논문에 적용한 나이브 베이즈 모델은 단어들이 서로 독립되어 있다고 전제하고, 특정 단어들을 포함하는 문서가 어느 주제에 속할지를 결정하는 방법이다. 이를 통해 특정 주제에 속할 확률이 가장



<그림 4> TF-IDF 분석과 나이브베이즈 분석 방법

큰 주제로 분류한다. 베이즈 정리(Bayes theorem)를 이용하여 개발된 나이브 베이즈 모델은 자료량이 많을수록 정확도가 높은 특징이 있다.

3) 감성지수 산출

본 연구에서는 방송뉴스 기사를 활용한 감성 분석(sentiment analysis)을 실시하여 부동산 감성지수를 개발한다. 감성분석은 긍정과 부정의 요소를 판별하고, 이를 추출, 범주화 및 분류, 해석하여 정량화하는 작업이다(남길임 · 조은경, 2017). 본 연구는 통계적 방법과 기계학습 모델을 접목한 감성사전의 구축과 감성지수의 산출을 실시하였다.

부동산 감성지수의 산출을 위해 우선, 1개 문장에 대한 긍정과 부정 감성의 가중치를 계산한다. 방송뉴스는 ‘긍정’과 ‘부정’의 극성을 가지고 있는 단어로 이루어진 문장의 집합으로 이루어져 있다. 첫 번째 단계는 나이브 베이즈 분류기법을 활용하여 각 문장을 감성사전과 비교하고, 각 문장에 대해 일별 가중치 합을 구한다.

다음 단계에서는 앞서 일별로 산출한 방송뉴스 문장의 긍정과 부정 가중치 합을 해당 월의 말일까지로 확대한다. 즉, 해당 월의 모든 문장의 긍정과 부정의 가중치를 해당 월의 날짜로 나누어 월단위 가중치 평균값 다음과 같이 구한다.

$$\begin{aligned}
 M.TVSI_p &= \frac{\sum_{i=1}^n D.TVSI_p(i)}{n} \\
 M.TVSI_n &= \frac{\sum_{i=1}^n D.TVSI_n(i)}{n}
 \end{aligned}
 \tag{식 1}$$

마지막 단계에서는 앞서 산출한 월별 긍정 가중치 평균값과 부정 가중치 평균값의 차이를 구한다. 긍정 가중치 평균값에서 부정 가중치 평균값을 차감하여 양(+)의 값이 나타나면 해당 월은 부동산 시장에 대해 긍정적인 감성을 가지고 있는 것으로 판단한다. 식은 아래와 같다.

$$TVSI_i = \sum_{i=1}^n M.TVSI_p(i) - \sum_{i=1}^n M.TVSI_n(i)
 \tag{식 2}$$

IV. 분석 결과

1. 기술 통계분석

1) 방송뉴스 감성지수

부동산 관련 방송뉴스 기사를 이용한 방송 감성지수는 뉴스기사의 수집 및 전처리, 토픽 분석을 이용한 토픽과 중심 단어 추출, 텍스트 랭크 알고리즘을 적용한 문장내 단어간의 관계 도출 및 감성사전 구축, TF-IDF를 활용한 뉴스기사의 문장에 대한 단어의 점수 계산, 나이브 베이즈 분류 모델을 활용한 문장의 긍정 및 부정 극성 부여 및 가중치 산정의 순서로 이루어진다. 이렇게 도출된 가중치를 감성지수 공식에 따라 월별 방송 감성지수를 도출한다.

기계학습 방법을 적용한 방송 감성지수의 시계열 그래프는 <그림 5>와 같다. 방송 감성지수는 기준값인 0을 중심으로 긍정 값이 높으면 위쪽에 감성지수가 위치하고, 부정 값이 높으면 0선 아래쪽에 감성지수가 위치한다. 분석기

간인 2012년부터 2013년 3분기까지 부동산 관련 방송 감성지수는 긍정과 부정으로 등락이 반복되었다. 2013년 8월 이후에는 2014년 2분기, 2017년 1분기와 2분기, 그리고 2018년 2분기를 제외하고, 대부분 긍정적 지수를 나타내고 있다.

방송 감성지수의 기초 통계량은 <표 2>와 같다. 평균값과 중앙값 모두 0보다 큰 이유는 2012년부터 2018년까지의 분석기간 중 부정적인 뉴스보다 긍정적인 측면을 부각한 뉴스가 많았기 때문이다.

산출한 감성지수의 정확성을 평가하기 위해 나이브 베이즈 분석의 정확도를 ROC(receiver operation characteristic) 곡선으로 평가한 결과, 정확도는 약 88%로 나타났다. 이는 감성지수의 정확성이 높은 수준임을 의미한다.

2) 서울시 아파트 매매가격지수

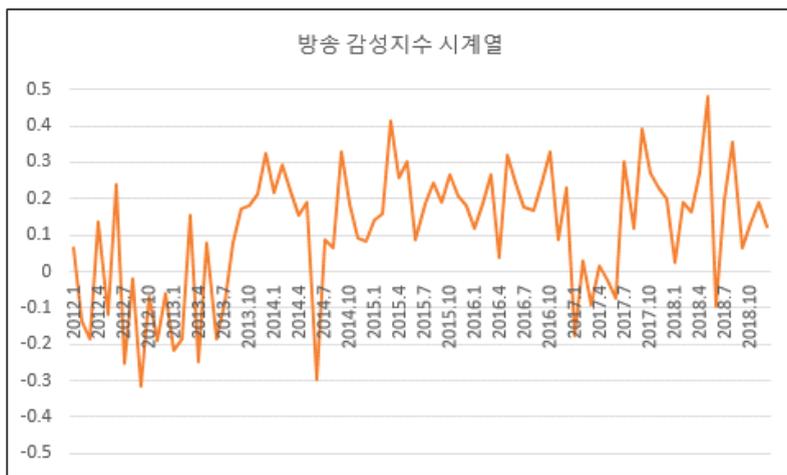
KB 국민은행의 2019년 1월 가격을 100으로

하여 산출한 서울 아파트 매매가격지수의 기초 통계량과 아파트 매매가격지수는 <표 3>, <그림 6>과 같다. 서울시 아파트 매매가격지수는 시계열자료로서 추세, 계절성, 주기, 불규칙성 등과 같은 성분들이 있다. 계절성을 제거하기 위해서 R통계 프로그램의 추세 분해를 이용하여 원자료에서 추세요인(trend), 계절성 요인(seasonal), 불규칙요인(random)을 분리한 후 계절요인 값을 가감하여 계절조정된 지수를 구하였다. 분석기간 서울시 아파트 매매가격지수는 2012년 1월부터 2013년 7월까지 하락하여 저점을 형성하고, 201년 1월부터 상승하기 시작하여 2018년 말까지 지속적으로 상승하였다.

2. 교차상관 및 인과관계 분석

1) 단위근 검정

KB 국민은행의 2012년 1월부터 2018년 12월까지의 서울시 아파트 매매가격지수와 방송



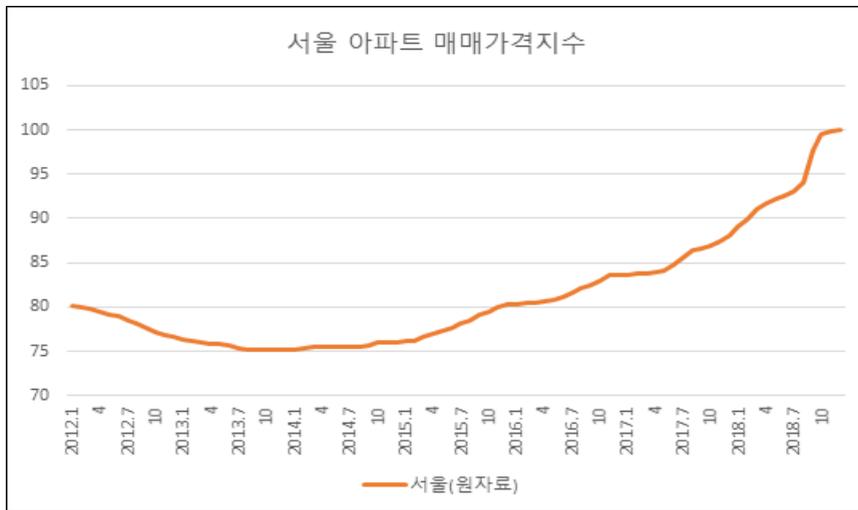
<그림 5> 방송 감성지수 시계열 그래프

〈표 2〉 방송 감성지수 기초 통계량

구분	평균	중앙값	표준편차	최소	최대
원자료	0.1112	0.1559	0.1760	-0.3151	0.4794
계절조정	0.1113	0.1294	0.1686	-0.2814	0.4320

〈표 3〉 서울 아파트 매매가격지수 기초 통계량(2019.1.=100)

구분	평균	중앙값	표준편차	최소	최대
원자료	81.31	79.45	6.469	75.14	100.14
계절조정	81.31	79.46	6.461	75.07	99.94



〈그림 6〉 서울 아파트 매매가격지수(2019.1.=100)

감성지수를 계절조정하고, 조정된 데이터를 이용하여 단위근 검정을 시행하였다. 단위근 검정은 ADF(augmented dickey-fuller) 방법을 이용하고, 유의수준은 5%를 적용하였으며, 절편항만 추가하였다.

분석 결과, 〈표 4〉와 같이 서울시 아파트 매매가격지수는 5% 유의수준에서 단위근을 갖는

다는 귀무가설을 기각하지 못하였다. 시계열적 안정성을 확보하기 위해서 서울 아파트 매매가격지수를 1차 차분하여 단위근이 없는 정상 시계열로 변환하였다. 방송 감성지수는 귀무가설을 기각하여 수준변수가 단위근을 가지고 있지 않는다. 따라서 방송 감성지수는 수준변수를 사용하였다.

2) 교차상관관계 분석

방송 감성지수와 서울시 아파트 매매가격지수 간 교차상관관계 분석 결과는 <표 5>와 같다. 분석 결과, 시차 4에서 가장 큰 값을 가지며 양의 상관관계를 맺고 있는 것으로 나타났다. 방송 감성지수가 서울시 아파트 매매가격지수에 선행하며, 4개월의 시차에서 상관관계가 가장 강하게 나타난다. 조태진(2014)은 부동산경기전망지수가 시차 5에서 주택가격에 가장 높은 상관관계를 보인다고 주장하였다.

시차 0을 중심으로 통계적으로 유의미한 상관관계수 값이 <그림 7>과 같다. 방송 감성지수의 증가(감소)가 서울시 아파트 매매가격지수의 증가(감소)를 유의미하게 이끄는 동시에 서울시 아파트 매매가격지수의 증가(감소)가 방송 감성지수의 증가(감소)에도 영향을 미친다. 아파트 매매가격의 상승 또는 하락 초기에 관련 뉴스

를 방송사들이 사전에 기사화하기 때문이라고 판단된다. 방송 감성지수가 선행하는 상관관계가 대부분 시차에서 후행하는 상관관계보다 크다. 감성지수의 선행 상관관계는 4개월 시차에서 가장 높은 반면, 후행 상관관계는 1개월 시차에서 가장 높다.

3) 그랜저 인과관계 검정

방송 감성지수와 서울시 아파트 매매가격지수 간의 인과관계를 검증하기 위해 최대 시차를 5개월로 설정하고, 그랜저 인과관계 검정을 실시하였다. 분석 결과, <표 6>과 같이 방송 감성지수는 서울시 아파트 매매가격지수에 시차 1부터 시차 5까지 유의미한 인과관계를 형성하는 것으로 나타났다. 방송뉴스를 통해 전달되는 부동산 감성지수가 서울시 아파트 매매가격지수의 등락에 유의미한 인과관계를 형성하며,

<표 4> 단위근 검정 결과

변수명	수준변수		1차 차분변수	
	t-statistics	p-value	t-statistics	p-value
서울 매매가격지수	3.6077	1.0000	-4.0204	0.002***
방송 감성지수	-3.0119	0.0379**	-	-

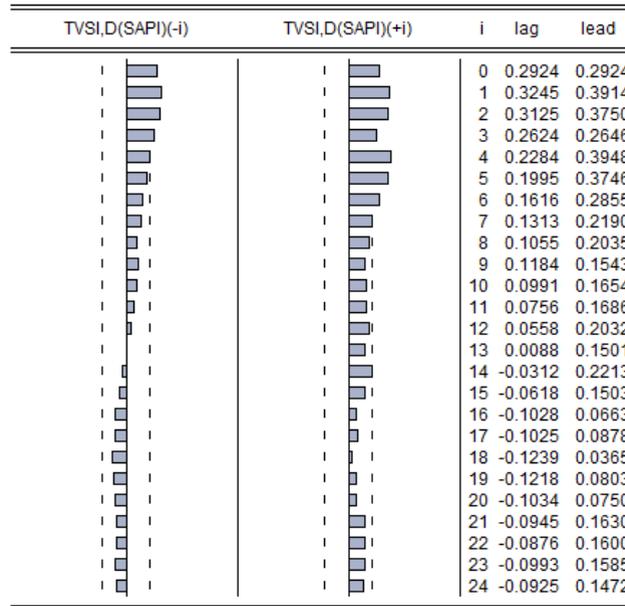
* p<0.10, ** p<0.05, *** p<0.01.

<표 5> 방송 감성지수와 아파트 매매가격지수의 교차상관관계 분석

시차	0	1	2	3	4	5
선행	0.2924	0.3914	0.3750	0.2646	0.3948*	0.3746
후행		0.3245	0.3125	0.2624	0.2284	0.19952

주 1: *는 교차상관계수 절대값이 가장 큰 것과 그에 상응하는 시차.

주 2: 1차 차분한 서울 아파트 매매가격지수를 사용.



〈그림 7〉 방송 감성지수와 아파트 매매가격지수의 교차상관 그래프

〈표 6〉 방송 감성지수와 아파트 매매가격지수의 그랜저 인과관계 검정

귀무가설	Lag 1	Lag 2	Lag 3	Lag 4	Lag 5
	F-통계량 (p-value)	F-통계량 (p-value)	F-통계량 (p-value)	F-통계량 (p-value)	F-통계량 (p-value)
방송감성지수 ⇒ 매매가격지수	6.5666** (0.0123)	4.4212** (0.0153)	2.5240* (0.0642)	3.7772*** (0.0077)	2.2401* (0.0603)
매매가격지수 ⇒ 방송감성지수	5.1655** (0.0258)	1.0802 (0.3447)	0.7401 (0.5315)	0.8237 (0.5145)	0.6580 (0.6565)

주 1: *, **, ***는 각각 유의수준 10%, 5%, 1% 이하에서 귀무가설을 기각하므로, 변수 상호 간에 그랜저 인과관계가 있음을 의미.

이 관계는 1개월부터 5개월에 이르기까지 유의미하게 작용하는 것으로 해석된다.

그러나 서울시 아파트 매매가격지수는 방송 감성지수에 대해 시차 1에서만 통계적으로 유의미한 그랜저 인과관계를 형성하는 것으로 분석되었다. 서울시 아파트 매매가격지수의 등락

이 방송뉴스를 통해 즉각적으로 긍정 또는 부정적 뉴스로 전달되어 긍정적 또는 부정적 부동산 감성지수에 인과관계를 형성하는 것으로 해석된다. 서울시 아파트 시장에서 방송 감성지수와 매매가격지수는 양방향의 인과관계를 형성하는 것으로 분석되며, 시간이 흐르면서

방송 감성지수가 매매가격지수에 일방향으로 영향을 미치는 것으로 볼 수 있다.

유한수 · 정재호(2015)는 소비심리지수와 주택매매가격지수가 양방향의 인과관계를 나타낸다고 주장하였다. 임재만 · 임미화(2017)도 소비심리지수와 뉴스는 주택가격지수에 상호 영향을 미친다고 주장하였다. 박재수 · 이재수(2019) 또한 부동산 감성지수와 서울시 아파트 매매가격지수가 전반적으로 양방향의 상관 및 인과관계를 나타낸다고 분석하였다.

V. 결론 및 함의

이 연구는 부동산 관련 방송뉴스 스크립트를 수집 · 분류하고 기계학습 기법을 통해 산출한 부동산 감성지수가 서울시 아파트 매매가격지수와 유의미한 상관 및 인과관계가 있는지 밝히는데 목적이 있다.

분석 결과와 시사점을 정리하면, 첫째, 2012년부터 2013년 3분기까지 부동산 관련 방송 감성지수는 긍정과 부정이 반복되었다. 2013년 8월 이후에는 2014년, 2017년, 2018년 2분기 등에서 부정 지수가 나타난 반면, 대부분 시기에 긍정 지수를 보인다.

둘째, 부동산 관련 방송 감성지수는 서울시 아파트 매매가격지수와 양의 상관관계를 나타낸다. 방송 감성지수의 증가가 서울시 아파트 매매가격지수의 증가를 이끌고, 매매가격지수의 증가 또한 방송 감성지수의 증가에도 유의미한 영향을 미친다. 감성지수가 매매가격지수에 선행하는 관계는

4개월 시차에서, 후행 관계는 1개월 시차에서 상관성이 높다.

셋째, 방송 감성지수와 서울시 아파트 매매가격지수 사이의 인과관계를 분석한 결과, 방송 감성지수는 매매가격지수에 1개월 시차부터 5개월 시차까지 모두 유의미한 그랜저 인과관계를 보이는 것으로 분석되었다. 방송뉴스를 통해 전달되는 감성지수가 서울시 아파트 매매가격지수의 등락에 유의미한 영향을 미치는 것으로 해석된다.

방송뉴스를 통해 전달되는 감성지수가 서울시 아파트 매매가격지수에 선행관계와 인과관계를 나타내는 것은 텍스트 마이닝과 비정형 빅데이터를 활용하여 부동산시장의 메커니즘을 설명 및 예측할 수 있음을 시사한다. 기계학습 등 인공지능을 활용하여 비정형 빅데이터로부터 산출한 감성지수 등이 부동산 부문 및 산업에 적용될 수 있다는 점도 시사한다.

부동산 관련 방송뉴스가 실제 부동산 가격에 영향을 미친다는 사실은 부동산 뉴스가 광고를 위해 작성되거나 정확한 보도를 위한 노력없이 부정확한 보도를 하게 되면 부동산시장에 악영향을 미칠 수 있다는 점을 의미한다. 따라서 방송 등의 언론은 부동산 뉴스에 대하여 신중한 사실 확인이 필요하고 사실에 기반한 정확한 보도를 하여야 한다. 무분별하고 부정확한 뉴스는 부동산시장을 왜곡시킬 수 있기 때문이다.

텍스트 마이닝 또는 빅데이터를 활용한 다양한 지수 산출에 관한 연구가 필요하다. 또한 이 지수가 가격뿐만 아니라, 부동산시장의 다양한 부문을 설명 및 예측하는데 유용한 지에 관한 후속 연구도 요구된다.

참고문헌

- 국토연구원. (2020). *부동산시장 심리지수*. Retrieved from http://kremap.krihs.re.kr/menu2/Mind_Land, 2020.10.22.
- 김대원, 유정석. (2013). 주택가격에 대한 심리적 태도가 주택매매 거래량에 미치는 영향 분석. *주택연구*, 21(2), 73-92.
- 김대원, 유정석. (2016). 트위터 정보와 아파트 매매 및 전세 가격 간 동적 관계 분석. *도시행정학보*, 29(1), 1-33.
- 김리영, 안지아. (2013). 소비자의 주택가치 전망이 가격 및 거래에 미치는 영향. *국토계획*, 48(3), 403-417.
- 김진유. (2006). 신문기사가 부동산가격변동에 미치는 영향: '투기'가 포함된 신문기사와 주택가격간의 그랜저인과관계분석을 중심으로. *주택연구*, 14(2), 39-63.
- 남길임, 조은경. (2017). *한국어 텍스트 감성 분석*. 서울: 커뮤니케이션북스.
- 노민지, 유선종. (2016). 소비자 심리와 아파트 실거래가격 간 관계분석: 인터넷 검색량 및 국토연구원 주택매매시장 소비심리지수를 중심으로. *국토연구*, 89, 3-13.
- 박재수, 이재수. (2019). 아파트 매매가격과 부동산 온라인 뉴스의 교차상관관계와 인과관계 분석: 온라인 뉴스기사의 비정형 빅데이터를 활용한 감성분석 기법의 적용. *국토계획*, 54(1), 131-147.
- 배영준, 장호택, 홍태원, 이해연. (2018). 향상된 TextRank 알고리즘을 이용한 자동 회의록 생성 시스템. *한국정보전자통신기술학회논문지*, 11(5), 467-474.
- 안정욱, 이규현, 김희웅. (2015). 정보시스템 연구 트렌드 변화 분석: 토픽모델링과 네트워크 분석. *한국경영정보학회 학술대회논문집*, 2015(11), 561-570.
- 여론집중도조사위원회. (2018). *뉴스 이용집중도 조사보고서 2016~2018*. 서울: 한국언론진흥재단. Retrieved from https://www.mcst.go.kr/kor/s_notice/press/pressView.jsp?pSeq=17050
- 우윤석, 이은정. (2011). 언론보도와 시계열 주택가격 간의 관계에 관한 연구. *주택연구*, 19(4), 111-134.
- 유한수, 정재호. (2015). 주택시장에서의 매매가격지수와 소비심리지수의 관계. *부동산연구*, 25(4), 49-61.
- 임재만, 임미화. (2017). 주택시장 뉴스, 소비심리, 가격, 거래량의 관계. *주택연구*, 25(3), 5-24.
- 조은경. (2015). 감성분석에서 함수적 어휘와 대하 여성 추출. *한국어 의미학*, 47, 1-26.
- 조태진. (2014). 심리지수가 주택시장에 미치는 영향에 관한 연구. *주택연구*, 22(3), 25-48.
- 진창하, Paul, G. M. (2012). 신문기사 내용과 주택가격: 인식, 사유, 그리고 투자심리. *부동산학연구*, 18(2), 125-142.
- 차윤정, 이지혜, 최지은, 김희웅. (2015). 소셜미디어 토픽모델링을 통한 스마트폰 마케팅 전략 수립 지원. *지식경영연구*, 16(4), 69-87.
- 한국언론진흥재단 미디어연구센터. (2015). *온라인 설문조사*(2015.9.19-23.). Retrieved from <https://www.kpf.or.kr/front/mediaStats/mediaStatsListPage.do>
- Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, 55(4), 77-84.
- Gayer, T., & Visicusi, W. K. (2002). Housing price responses to newspaper publicity of hazardous waste sites. *Resource and Energy Economics*, 24(1-2), 33-51.
- Lee, S. G., Lee, B. S., Bark, B. Y., & Hwang, H. K. (2010). A study of intelligent recommendation

system based on Naïve Bayes text classification and collaborative filtering. *Journal of Information Management*, 41(4), 227-249.

Stevens, D., Wubben, S., & van Zaanen, M. M. (2014). *Predicting real estate price using text mining* (HAIT Master Thesis). Tilburg University School of Humanities, Tilburg, Netherlands.

Sun, D., Zhang, C., Xu, W., Zou, M., Zhou, J., & Du, Y. (2014). Does web news media have

opinion: Evidence from real estate market prediction. *Proceedings of the 18th Pacific Asia Conference on Information Systems* (No. 374).

논문접수일: 2020.10.22

논문심사일: 2020.11.11

게재확정일: 2020.11.19

Journal of Housing and Urban Finance 2020; 5(2):53-68
pISSN: 2508-3872 | eISSN: 2733-4139
<https://doi.org/10.38100/jhuf.2020.5.2.73>

The correlation and causality between broadcasting sentiment index and housing sales price index: Evidence from Seoul

Jae-Su Lee*, Jae-Soo Park**

Abstract

The study intends to understand whether the real estate sentiment index, calculated using real estate-related broadcasting news, has a significant correlation or causal relationship with the apartment sales price index based on data from 2012 to 2018 in the Seoul Metropolitan Area. We find that the broadcasting sentiment index correlates positively with the apartment sales price index; an increase in the broadcasting sentiment index results in an increase in the sales price index in Seoul, which again has a significant impact on the increase in the broadcasting sentiment index. In addition, the broadcasting sentiment index indicates a significant causal relationship between one-month and five-month time difference with the sales price index. The broadcasting sentiment index has a significant impact on the fluctuations in the apartment sales price index. It suggests that the mechanism of the real estate market can be explained and predicted by utilizing text mining and unstructured big data. It also implies that artificial intelligence techniques such as machine learning can be applied to the field of real estate.

Key words: housing sales price, broadcasting news, sentiment index, cross correlation, Granger causality

* Associate Professor, Department of Real Estate, Kangwon National University, E-mail: jslee25@kangwon.ac.kr

** Executive Director, Hanil Networks Co., Ltd., E-mail: okparkjaesoo@naver.com

© Copyright 2020 Korea Housing & Urban Guarantee Corporation. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.