



텍스트 마이닝을 통한 관광지 수요예측*

- 온라인 검색 엔진을 중심으로 -

Forecasting Tourism Demand using Text Mining Techniques: Focused on an Online Search Engine

박수지** · 신진옥*** · 송상헌**** · 정철*****

Park, Su-Ji · Shin, Jin-Ok · Song, Sang-Hun · Jeong, Chul

요약 : 본 연구는 온라인 검색 엔진을 활용한 관광지 수요예측을 실시하기 위하여 온라인 검색 엔진에서 도출된 텍스트 마이닝 기반 이슈 키워드 분석 자료를 활용하였다. 구체적으로 첫째, 관광정보시스템 관광객 수 추이와 텍스트 마이닝을 통해 도출된 이슈 키워드 검색 관광객 수 추이와의 유의성 판단을 위해 두 데이터를 표준화하여 그 수치를 비교하였다. 둘째, 온라인 검색 엔진을 통해 도출된 데이터와 관광지식정보시스템 통계 데이터를 활용한 수요예측을 실시한 후 그 추이를 비교하였으며, 실제 관광객 수로 변환하였다. 셋째, 이를 토대로 이론적 및 실무적 시사점을 제시하였다. 그 결과, 온라인 검색 엔진을 활용한 수요예측 시 안동시 관광객 수는 2018년도 까지 지속적으로 증가하는 것으로 나타났으나, 관광지식정보시스템을 통한 수요예측 결과에서는 관광객 수가 감소하는 것으로 나타났다. 이러한 결과를 토대로 실무적 시사점을 제시하였다.

핵심용어 : 텍스트 마이닝, 수요예측, 온라인 검색 엔진, 빅데이터, 관광정보탐색행동, 안동

ABSTRACT : The purpose of this paper is to prove demand forecasting methods by utilizing online search engines. This study utilizes text mining based issues and Keyword analysis data derived from online search engines to predict the number of visitors that exist demanding forecasting. In terms of forecasting tourists, this study compared NAVER trend statistics to the statistics of a Tourism Information System. and, it was carried out by applying the most appropriate model to forecast demand time series data. Also, it compared existing tourist forecast statistic results of a Tourism Information System in order to forecast the result of statistical NAVER trends. As a result, we have found that when utilizing NAVER trends the forecast is for Andong tourists to continue to increase until 2018 year. However, forecasting results though the Tourist Information System showed that the number of tourists decreased. This study proposes practical implications based on these results.

Key words : Text mining, Demand forecasting, On-line search engine, Big-data, Tourist information search behavior, An-Dong

* 이 논문은 2015년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (2015R1A5A7037751).

** 한양대학교 대학원 관광학과 석사과정. e-mail: berrypsj@hanyang.ac.kr

*** 한양대학교 관광학부 BK21+ 스마트관광 창조인력양성 사업단 산학협력전담연구원. e-mail: jo1477@hanyang.ac.kr

**** 하얏트리젬시 제주. e-mail: akon7@hanmail.net

***** 한양대학교 관광학부 조교수(교신저자). e-mail: jeong72@hanyang.ac.kr

I. 서 론

정보통신 기술의 급속한 발전은 정보 공유 확산과 함께 정보탐색활동에 적극 참여할 수 있는 환경을 조성하였다(서정호·김철원, 2012; Taylor, Lewin & Strutton, 2011). 관광객 정보탐색 노력에 있어서 인터넷이 보다 중요시 되는 이유는 인터넷을 통해 획득한 정보 비중이 높아지고 있기 때문이다(정철·이준남, 2010; Wang & Fesenmaier, 2006). 관광객은 정보탐색을 통해 얻게 된 정보로 목적지에 대한 이미지를 체계화 시키며 최종적으로 관광지를 선택하게 된다(표원정·임관혁, 2012). 관광객 의사결정은 정보탐색, 수집, 평가 등의 과정을 거쳐 최종의사결정을 하기 때문에 정보탐색과 관광지 선택은 직접적인 연관관계가 있다. 다양하고 방대한 양의 정보를 가진 온라인 검색 엔진은 정보탐색 행동 시 가장 많이 사용되고 있는 도구 중 하나로 그 대표적인 사이트는 구글, 네이버 등이 있다. 이러한 온라인 검색 엔진을 통해 얻게 되는 방대한 양의 정보를 빅데이터(big data)라 하는데 빅데이터는 단순히 큰 데이터라고도 하며 속성이 다양하고 디테일한 정보(안창원·황승구, 2012)를 수집하여 보다 쉽게 유용한 연구결과를 도출해 낼 수 있다는 장점을 지니고 있어(Kitchin, 2014) 다양한 연구 분야에서 사용하고 있다(조완섭·조아·권가은·류관희, 2015; 한진성·윤지환, 2016; Lau, Lee & Ho, 2005; Loh, Lorenzi, Saldaña & Licthnow, 2003).

관광분야에서의 수요예측은 관광개발 및 관광정책 수립 시 기초자료로 활용되는데 이는 불확실한 미래상황을 추정하고 예측할 수 있기 때문에 매우 중요한 연구 분야이다(손은호·서진우·정명보, 2005; 송근석·이혜숙, 2006; 최병길·오상훈·현정석, 2006). 관광분야에서는 수요예측 시 '관광지식정보시스템' 제공 통계 및 지자체에서 집계하는 통계 등의 로우 데이터(raw

data)를 활용하여 수요예측 분석을 실시하고 있다. 그 중 가장 많이 사용되는 데이터는 관광지 지식정보시스템 제공 통계이다. 관광지 지식정보시스템은 2013년 이전 무료관광지와 유료관광지로 구분하여 측정하였는데 유료관광지는 실제 관광객 수 측정을 통한 비교적 정확한 데이터를 제공하는데 반해 무료관광지는 추정치로 관광객 과대 집계라는 논란이 발생하여 2013년 이후 부터는 특정관광지에 대한 측정만 이루어지고 있는 실정이다. 따라서 정부기관 및 지자체에서 제공하는 통계 데이터를 활용한 관광객 수요예측에서 벗어나 새로운 방식의 수요예측이 필요한 실정이다. 타 분야에서는 기존 통계 데이터를 활용한 수요예측 방식에서 벗어나 구글 트렌드, 네이버 트렌드 등의 온라인 검색 엔진을 통해 얻은 빅데이터를 수요예측에 사용하는 연구가 진행되고 있다(Ginsberg, Mohebbi, Patel, Brammer, Smolinski & Brilliant, 2009; Heerschap, Ortega, Priem & Offermans, 2014; Preis, Moat & Stanley, 2013). 이에 관광분야에서도 빅데이터를 활용한 새로운 방식의 수요예측이 필요하다고 판단된다.

텍스트 마이닝(text mining)은 빅데이터 정보를 분석할 수 있는 중요한 기법으로 주목되고 있다(심영석·김홍범, 2016). 텍스트 마이닝은 비구조적인 데이터를 대상으로 텍스트 간 암묵적인 정보를 추출하는 과정으로 비구조적 텍스트 데이터들을 분류하고 문서들 간 연관관계를 파악할 수 있는 방법이다(김휴찬·김미정·신현정, 2014). 이러한 텍스트 마이닝은 방대한 양의 온라인 관광정보를 분석하기에 적합하여 지역관광 마케팅 전략 도출(심영석·김홍범, 2016; 조완섭 외, 2015; 한진성·윤지환, 2016)을 위한 연구에 활용되고 있으나 텍스트 마이닝과 수요예측을 연계한 분석 방법은 미비한 실정이다. 문화체육관광부(2014)는 온라인 검색 엔진을 통한 텍스트 마이닝을 실시하여 국립공원 수요예측을

실시한 바 있다. 따라서 이러한 선행연구를 바탕으로 관광객 정보탐색행동과 관광지 선택행동의 연관관계가 발생할 것이라 판단하여 텍스트 마이닝을 통한 수요예측을 실시하고자 한다.

본 연구는 관광개발 및 관광정책 수립 시 중요한 자료로 활용되고 있는 수요예측이 보다 정확한 분석이 이루어져야 한다고 판단되어 이에 타분야에서 온라인 검색 엔진을 활용해 수집한 빅데이터를 수요예측에 적용하는 연구를 시도하고자 한다. 이를 위해 온라인 검색 엔진을 활용한 텍스트 마이닝 기반 이슈 키워드 분석을 실시하여 관광업계에 실질적인 수요예측 정보를 제공하고자 한다. 본 연구의 구체적인 목적은 다음과 같다. 첫째, 기존 관광지식정보시스템의 관광객 수 추이와 텍스트 마이닝을 통해 도출된 이슈 키워드 검색 추이와의 유의성 판단을 위해 두 데이터의 표준화 수치를 비교하고자 한다. 둘째, 네이버 트렌드를 통해 도출된 데이터와 관광지식정보시스템 통계 자료를 활용한 수요예측 추이를 비교하며, 실제 값으로 변환하고자 한다. 셋째, 이를 토대로 이론적 및 실무적 시사점을 제시하고자 한다.

II. 이론적 고찰

1. 온라인검색 엔진

관광객은 관광을 하고자 하는 욕구를 가지게 되면 그 욕구를 충족시키고자 대안을 탐색, 평가하여 최종적인 의사결정을 내리는데 이를 관광객 의사결정과정으로 설명할 수 있다(신진옥·정철, 2016; 이시우, 2002). Mathieson and Wall(1982)은 관광객 의사결정과정을 관광욕구 단계, 정보탐색과 평가 단계, 관광의사결정단계, 여행준비와 관광경험단계, 평가단계 등 5단계로 구분하였다. 관광객 의사결정과정에서 관광

객이 얻는 적절한 정보는 의사결정 시 많은 영향을 받기 때문에 정보탐색행동은 중요한 역할을 한다. 관광객은 목적지에 대한 이미지를 체계화시키기 위해 정보를 탐색하고 수집·평가하여 최종 의사결정을 내리게 되므로 관광객의 정보탐색행동은 관광 목적지 선택행동을 위한 의사결정 시 가장 중요한 단계라 할 수 있다(정철, 2009; 표원정·임관혁, 2012).

관광정보탐색행동은 관광객이 관광의사결정을 하기 바로 전 단계로 구체적이고 효과적인 관광활동을 위해 다양한 관광정보를 탐색하는 행동이다(황현철·박재현, 2000). 관광객 의사결정에서의 정보탐색행동은 관광계획 시 인터넷을 통한 정보탐색이 관광행동에 많은 영향을 미침을 밝히고자 하는 다양한 연구가 진행되었다(서정호·김철원, 2012; 전창석, 2003; 정병욱, 2013; Fodness & Murray, 1999; Gursoy & McCleary, 2004; Xiang & Gretzel, 2010). Xiang and Gretzel(2010)은 구글 검색엔진에서 관광정보탐색 시 어떠한 정보를 가장 많이 검색하는지를 밝히고자 하였다. 이를 위해 구글 검색엔진 검색 결과를 통한 빅데이터 분석을 실시하였는데 그 결과 관광정보탐색 시 소셜 미디어가 가장 큰 영향을 미치며 소셜 미디어 중 가상 커뮤니티(40%), 댓글(27%), 블로그(15%), 소셜 네트워크(9%) 등이 가장 많이 활용하고 있는 것으로 나타났다. 서정호·김철원(2012)은 국내 관광객이 관광을 계획하는 단계에서 의사결정요인으로 소셜 미디어가 어떻게 적용되며 어떠한 관계성을 형성하는지를 밝히고자 하였다. 이를 위해 인터넷 검색엔진 검색결과를 활용하여 도출된 온라인 관광 도메인내용분석을 실시하여 소셜 미디어와 비소셜 미디어로 분류하였으며 관광관련 키워드별 소셜 미디어 분포를 파악한 결과 소셜 미디어가 온라인 관광 도메인에 있어 중요한 역할을 한다는 점을 발견하였다. 전창석(2003)은 외부정보원천별 정보탐색 수준이 구매결정과

어떤 관계가 있는지, 어떤 정보원이 정보탐색에 있어 구매결정에 보다 많은 영향력을 미치는지를 비교 분석하였다. 그 결과 구매결정에 가장 큰 영향을 미치는 정보탐색형태는 언론매체를 통한 광고, 신문, TV, 온라인 검색 등으로 나타났다. 또한 현용호·허향진·김운숙(2004)은 온라인 관광객 의사결정과정에서 인터넷 기능이 관광객 의사결정과정에 어떠한 역할을 수행하는지에 대해 탐색적 연구를 실시하였다. 관광객은 여러 관광 웹사이트들을 탐색하여 탐색활동을 통해 얻게 된 정보를 비교분석한 후 관광지를 결정한다고 하였다. 이러한 선행연구를 종합하면 온라인 검색 엔진을 통한 온라인 관광정보 탐색 행동은 관광객 의사결정에 직접적 영향을 미치는 것으로 판단할 수 있다.

2. 텍스트 마이닝

빅데이터는 거대하면서도 디테일한 정보를 갖고 있기 때문에 저렴한 비용으로도 유용한 연구 결과를 도출할 수 있다(Kitchin, 2014). 과거와 달리 빅데이터는 정제되어 제공되기 때문에 누구나 쉽게 데이터를 다룰 수 있어 세계적으로 많은 연구가 진행되고 있다(Ginsberg *et al.*, 2009; Preis *et al.*, 2013). 다양한 연구분야에서 빅데이터의 중요성은 점점 높아지고 있으며 특히 다양하고 복잡한 문제를 효과적으로 해결하기 위한 방법으로 사용되고 있다(Labrinidis & Jagadish, 2012). 빅데이터 분석은 시계열 분석, 클러스터 분석, 네트워크 분석, 데이터마이닝 등의 방법으로 연구가 시도되고 있는데 그 중 트위터, 블로그 등 소셜 미디어에 저장된 각종 텍스트에서 의미 있는 정보를 추출하는 텍스트 마이닝이 큰 주목을 받고 있다(조완섭 외, 2015; 한진성·윤지환, 2016; Lau *et al.*, 2005; Loh *et al.*, 2003). 텍스트 마이닝은 주로 기술예측, 마케팅, 생물학 등의 분야에서 활용되고 있다(김

성근·조혁준·강주영, 2016).

관광분야에서는 Lau *et al.*(2005)이 단순히 경쟁 호텔의 가격정보 보다 텍스트 마이닝을 통한 고객의 기호와 통계 등의 데이터를 분석하여 합리적인 가격을 도출해 낼 수 있다고 하였으며, 조완섭 외(2015)는 SNS상의 중복관광 관련 글을 수집하여 이슈분석, 연관분석, 감성분석 등을 실시하였다. 분석을 위해 주요 블로그, 뉴스, 관광관련 공공 데이터 등 온라인상에 흩어져 있는 데이터를 수집하였으며 분석결과를 토대로 홍보 전략 및 정책 수립 방안을 제시하였다. 심영석·김홍범(2016)은 관광정보를 공유하는 대표적인 온라인 여행 커뮤니티 리뷰를 대상으로 텍스트 마이닝 기법을 이용하여 관광지 이미지의 구성요인을 밝히고 수집된 데이터를 통해 관광지 유형에 따른 관광지 이미지를 측정하였다. 안명숙·민용기(2016)는 호텔 이용에 대한 동기를 파악하고자 빅데이터를 활용하여 관광동기 이론을 바탕으로 고객들의 push 요인과 pull 요인을 텍스트 마이닝 하여 인터넷이 호텔 이용 동기에 큰 영향을 미치는 것을 밝혔다. 한진성·윤지환(2016)은 소셜 미디어 빅데이터를 텍스트 마이닝 하여 부산국제영화제에 대한 대중의 관심과 감정정도를 개최 전·후로 비교하였다.

이와 같은 선행연구를 바탕으로 텍스트 마이닝을 방대한 양의 정보에서 어떤 특정의 것에 대한 의미 있는 텍스트를 도출하는 분석 기법으로 정의하고자 한다. 이에 본 연구에서는 온라인검색 엔진을 활용한 빅데이터 수집을 통해 텍스트 마이닝을 실시하여 의미 있는 키워드를 추출하는 텍스트 마이닝 분석을 적용하고자 한다.

3. 수요예측

수요예측은 정책 결정 및 관리, 사업의 타당성을 검토하는 중요한 역할(Ma, Liu, Li & Chen, 2016)을 하며, 관광분야에서도 관광개발 및 정책

수립 시 중요한 자료로 활용되고 있다(교통개발연구원, 1988; 허향진·송재호, 2000; 황경후·김원표·정철, 2015; Lee, Song & Mjelde, 2008). 이러한 관광수요예측은 크게 계량기법과 질적 기법이 있으며, 계량 기법에는 시계열 모델과 인과 모델로 세분화되고 질적 기법에는 델파이 모델 등이 있다(이충기, 2003). 수요예측 기법은 사용하는 데이터에 따라 그 기법도 달라질 수 있다(Lee *et al.*, 2008; Lim & McAleer, 2001).

관광분야에서의 수요예측은 대부분 계량기법 중 시계열 모델에 집중되어 있다. 허향진·송재호(2000)는 제주도 입도 관광객의 시계열 자료를 활용하여 추세분석법, 지수평활법, Box-Jenkins ARIMA 등 시계열 분석 모형들을 통해 2000년부터 2005년까지 제주지역 관광객 수요를 예측하였다. Burger, Dohnal, Kathrada and Law(2001)는 남아프리카 여행을 위한 미국인 관광객 수요를 예측하기 위하여 시계열 데이터를 기반으로 지수평활, ARIMA, 다중회귀 분석 등의 방법을 사용하여 수요를 예측하고 비교하였으며, 황경후 외(2015)는 독도관리사무소에서 집계되는 독도 입도객 데이터를 활용하여 ARIMA, 계절분해법, 단순계절모형 등 일반량 시계열 모형 분석을 통한 독도관광객 수요예측을 실시한 결과, 계절 ARIMA 모형이 독도 관광 수요예측을 실시하는데 가장 적합하였으며, 수요예측 결과에 따라 교통편의 포화된 수요를 극복하기 위한 방안을 제시하였다. Ma *et al.*(2016)은 호주에 방문하는 중국인 관광객 수요가 지속적으로 상승할 것으로 계절성 ARIMA 모델을 통해 예측하며 향후 숙박시설 공급이 필요하다고 제시하였다. 이러한 선행연구들을 살펴본 결과, 관광분야에서의 수요예측은 ARIMA 모델을 사용하고 있는 것으로 보여진다(Lee *et al.*, 2008; Lim & McAleer, 2002). 대부분의 국내 관광객 수요예측 연구에서는 공통적으로 지자체, 관광지 등에서 시계열 자료를 수집하여 활용하고 있다는

것을 선행연구를 통해 알 수 있다. 그러나 지역에서 집계한 통계정보는 누락, 중복, 과대추정 등의 다각적인 문제점을 안고 있어 그 신뢰도에 대한 의문이 제시된 바 있다(노컷뉴스, 2016. 10.13; 전북도민일보, 2013.12.16; 중부일보, 2014.04.21). 과거의 관측 값을 활용하는 시계열 분석에서 관측 값의 신뢰도 하락은 큰 문제점으로 제기될 수 있다.

그러나 최근 온라인 기술의 발전으로 타 분야에서는 통계 자료 외 빅데이터를 활용한 수요예측을 실시하고 있다(Ginsberg *et al.*, 2009; Preis *et al.*, 2013). Ginsberg(2009)는 구글 트렌드를 활용한 독감 증상 예측에 관한 연구를 진행하였다. 구글에서 '독감 검색 수'와 '실제 독감환자 수'의 관계가 비슷한 것으로 나타났으며 이러한 분석결과를 통해 독감 환자 발생을 예측하였다. Preis(2013)는 전자상거래 시장 수요 예측에 관한 연구를 진행하였는데 투자자 불안심리가 검색활동과 관계가 있음을 밝혔다. 투자자 불안심리가 증가할수록 검색활동의 증가가 나타나고 이전 몇 주간 평균 횡수에 비해 열람 혹은 편집 횡수가 증가할 경우 투자자들의 불안심리에 따른 매도 행진과 주가 하락을 예측할 수 있다고 연구하였다.

관광분야에서도 빅데이터를 활용한 수요예측이 실시되고 있다. Heerschap *et al.*(2014)은 스마트 폰을 통해 얻은 로우데이터를 활용하여 관광 통계를 도출하는 연구를 진행하여 국적별 관광객들이 어떤 지역에 많이 체류하고 오래 있는지 패턴을 분석하였으며, 이를 통해 스마트폰 로우데이터가 관광객의 수요를 예측하기에 적합하다는 것을 밝혔다. 또한 Artola, Pinto and de Pedraza Garcia(2015)는 스페인 여행과 관련된 구글 검색키워드를 통해 구글 트렌드 데이터를 분석하여 ARIMA 모델을 통한 관광객 수를 예측하고자 하였다. 그 결과 영국, 독일, 프랑스에서 스페인에 방문한 실제 방문객과 구글

트렌드 검색 추이가 유사하다는 것을 밝혔으며 ARIMA 모델을 활용한 수요예측은 단기예측 시 실효성이 있음을 밝혔다. 문화체육관광부(2014)는 온라인 검색 엔진을 활용한 국내외 관광객 수요예측을 실시하였는데 외래 관광객 수요예측을 위해 구글 트렌드를, 국내 관광객 수요예측을 위해 네이버 트렌드를 사용하였다. 외래 관광객 수요예측은 일본, 중국 등 주요 인바운드 관광객들의 국적별 순위에 따라 10개국을 선정하여 구글 검색 횟수를 데이터화한 수치를 변수로 사용하였다. 그리고 국내 수요예측은 국내 국립공원 수요예측을 실시하였는데 이를 위해 네이버 트렌드를 통한 관련 키워드를 도출하여 실제 방문객 수와의 추이 비교 및 검색 횟수를 사용해 국내 관광 수요예측방법이 기존 수요예측보다 더 예측력이 뛰어남을 검증하였다. 그러나 단순 검색추세만 비교하였을 뿐 실제 관광객 수를 도출해내지는 못하였다.

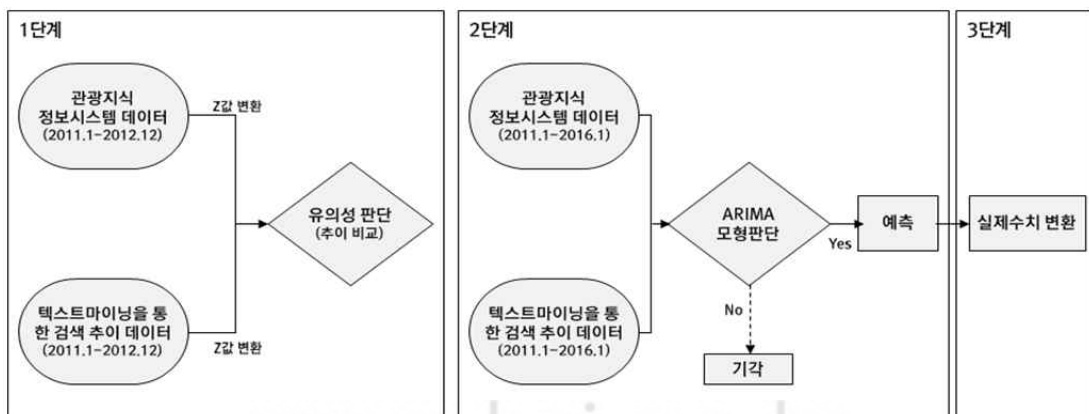
이와 같은 선행연구를 통해 온라인검색 엔진을 통한 수요예측은 기존 통계데이터를 활용한 수요예측보다 효율적이며 그 중 ARIMA 모형을 활용한 수요예측은 단기예측 시 그 실효성이 높을 것으로 판단할 수 있다. 따라서 본 연구에서는 온라인검색 엔진을 통해 수집된 데이터를 기반으로 텍스트 마이닝 분석을 실시한 후 도출된

키워드 추이 데이터를 활용한 수요예측을 실시하고자 한다.

Ⅲ. 연구방법

1. 연구진행 과정

본 연구는 온라인 검색엔진을 활용한 수요예측을 실시하고자 한다. 이에 본 연구의 진행과정은 구체적으로 <그림 1>과 같다. 1단계에서는 관광지식정보시스템의 주요관광지 관광객 수와 텍스트 마이닝을 통한 검색 추이 데이터를 표준화한 값을 순차도표를 통해 추이 비교를 실시하여 유의성을 판단한다. 2단계에서는 2011년 1월부터 2016년 1월까지의 관광지식정보시스템 데이터와 텍스트 마이닝을 통한 검색 추이 데이터를 확보하였다. 확보된 데이터는 계절성이 나타나기 때문에 계절성 분석이 가능한 ARIMA(p,d,q)(P,D,Q)₁₂에 적합한 모형을 판단하여 해당 모형을 통해 2018년까지의 검색 수치를 예측하였다. 3단계에서는 텍스트 마이닝을 통한 검색 추이 데이터의 특성상 검색횟수를 상대 값으로 0에서 100까지 숫자로 환산하였으므로 실제 관광객 수치로 변화시키기 위한 X값을 구하



<그림 1> 수요예측 진행과정

는 공식으로 적용하여 2018년 까지 방문할 것으로 예상되는 관광객의 수치를 변환하는 순서로 진행하였다.

2. 분석방법

본 연구는 SPSS 20.0을 활용하여 일변량 시계열 모델인 이동평균법, 지수평활모델, ARIMA 모델 중 Box-Jenkins ARIMA 모델을 사용하였다. 본 모델은 수요예측 방법을 할 때 유용한 방법이며 비교적 간단한 시계열 자료만 있어도 통계적으로 신뢰 있는 수요예측을 할 수 있는 장점을 갖고 있다(Lim & McAleer, 2002). 또한 예측치의 정확도가 높기 때문에 중·단기의 관광 수요예측에 적합한 방법이다(이충기, 2003).

3. 연구대상지 선정

본 연구는 한국의 정신문화 수도라 불리는 안동시를 연구대상지로 선정하였다. 안동시는 경상북도 북부 중심에 위치한 도시로 경상북도에서 가장 면적이 넓은 뿐만 아니라 도시 전역에 다양한 관광자원을 보유하고 있어 2011년부터 2014년까지 평균 450만 명 이상의 관광객이 찾고 있다(관광지식정보시스템, 2015). 특히 안동시는 세계문화유산인 하회마을을 비롯한 도산서원, 봉정사 등 다양한 유교문화자원을 지니고 있으며, 하회마을 권역, 도산서원 권역, 안동시내 권역 등 4가지 유형의 시티투어를 운영, 코레일 내일로 티켓 이용자를 위한 다양한 프로모션 진행(경북일보, 2016.01.12) 등 관광객 유치를 위해 많은 노력을 하고 있다. 이와 같이 한국의 대표적인 유교문화 도시인 안동시를 대상으로 분석함으로써 안동 관광 활성화에 기여할 수 있을 것으로 판단된다.

3. 자료수집

1) 관광지식정보시스템

관광지식정보시스템은 2002년부터 문화체육관광부의 국가 관광정보화 추진 전략 계획에 근거하여, 관광정보의 표준화 및 정보의 유통체계 확립을 위해 한국문화관광연구원에서 연차사업으로 추진, 운영되어 오고 있다. 이러한 정보시스템은 관광과 관련된 지식 및 정보 제공 뿐 아니라 인터넷 사이트를 통해 관광과 관련된 통계자료 등을 제공하고 있다. 본 연구에서는 각 지자체에서 주요 관광지별로 집계한 관광객 수를 통합하여 제공하고 있는 주요관광지점 입장객 통계 중 안동시 전체 데이터를 수집하여 분석 자료를 활용하였다.

2) 키워드 검색 추이

안동시 관광과 관련한 키워드 중 어떠한 키워드가 가장 많이 검색되고 있는지를 도출하고자 텍스트 마이닝을 수행하였다. 2016년 1월 27일 기준 '네이버 연관키워드조회'에서 가장 많은 조회 수를 기록하고 있는 상위 5개 키워드인 '안동 여행', '안동 펜션', '안동 가볼만한 곳', '안동 볼거리', '안동 맛집'으로 도출하였다.

3) 검색 횟수를 활용한 관측 값 확보

네이버 검색엔진에서는 기간별 검색 추이 및 키워드 관련 그래프를 제공하는 검색 통계 서비스인 네이버 트렌드를 운영하고 있다. 본 연구에서는 '네이버 연관키워드조회'에서 도출된 5개 키워드를 네이버 트렌드에 적용하여 관측 값을 확보하였다. 관측 값은 ARIMA 모형을 설정하기 위한 데이터로 이를 위해 2011년 1월부터 2016년 1월까지의 각 키워드별 검색 트래픽 수치 62개를 확보하였다. Box and Jenkins(1976)는 ARIMA 모형을 설정하기 위해서 최소 50개 이상의 관측 값이 필요하다고 언급하였는데 이에

본 연구에서는 62개의 관측 값을 확보하였기 때문에 분석 상 무리가 없을 것으로 판단된다.

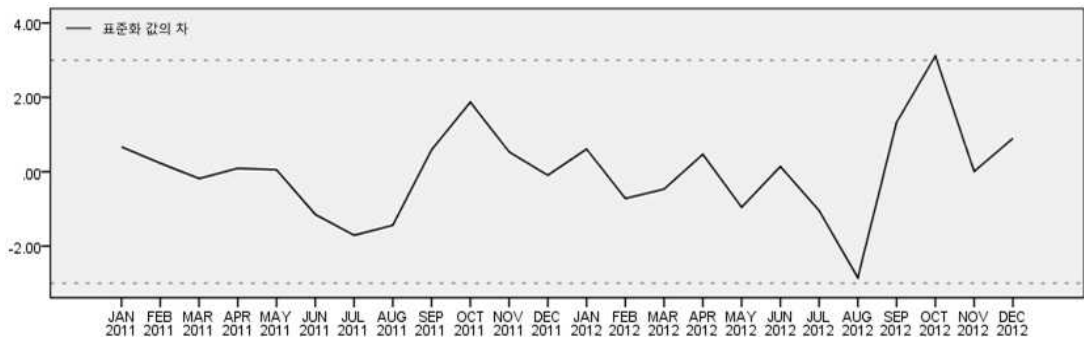
IV. 분석결과

1. 집계 관광객 수와 텍스트 마이닝 데이터와의 추이 비교

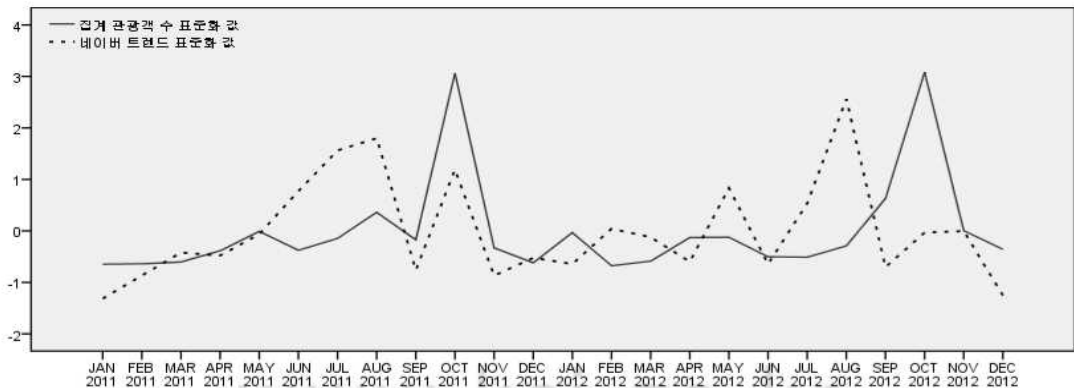
관광지식정보시스템에서 집계되고 있는 관광객 수와 네이버 검색 통계 값을 비교하여 어떠한 차이가 있는 지 검증하였다. 동일한 단위로 두 데이터를 비교하기 위하여 집계된 관광객 수와 텍스트 마이닝을 통해 도출된 값을 표준화시킨 값을 산출하였다. 두 데이터의 표준화 값 차를 구한 후 순차도표를 그린 결과는 <그림 2>와 같

다. 순차도표에 따른 결과 두 데이터는 ±3의 범위 안에서 유사한 추이를 보였다.

두 데이터의 개별 추이를 구체적으로 분석한 결과, <그림 3>과 같이 정보탐색이 집중적으로 이루어지는 시기와 관광지식정보시스템에서 관광객 수가 증가하는 시기가 일치하지 않는 것으로 나타났다. 이러한 결과의 원인은 두 가지로 분류할 수 있다. 첫째, 관광지식정보시스템이 한정된 관광자원에서 관광객을 집계하기 때문인 것으로 추정된다. 네이버 트렌드 검색 추이 결과 7~8월 관광객이 높게 나타났으나 관광지식정보시스템 집계 결과 낮은 추이를 나타냈다. 이는 7~8월에 휴가철 관광객 및 코레일 내일러(내일로 티켓을 이용하는 학생)가 급증(대구일보, 2016.03.21; 한겨레신문, 2015.08.21)하는데 이러한 관광객이 실제 방문한 관광지에서 관광객 수



<그림 2> 관광지식정보시스템과 네이버 트렌드 표준화 값의 차



<그림 3> 개별 추이 비교 분석

를 집계하지 않았기 때문으로 설명할 수 있다. 둘째, 관광지식정보시스템에서 중복집계가 되었기 때문으로 추정할 수 있다. 관광지식정보시스템의 관광객 수 추이를 살펴보면 안동국제탈춤페스티벌이 개최되는 10월 안동시 관광객의 약 20%가 집중 방문하는 것으로 나타났다. 그러나 안동국제탈춤페스티벌은 하회마을 및 안동시내 일대에서 개최되나 관광지식정보시스템에서는 하회마을과 안동국제탈춤페스티벌을 중복하여 집계하였다. 따라서 중복집계로 인한 과대추정이 이루어졌을 것이라 판단된다.

2. 텍스트 마이닝 데이터를 활용한 시계열 분석

1) 계절 ARIMA 최적화 모형 선정

시계열 분석을 위해 텍스트 마이닝을 통해 추출한 데이터를 다음과 같은 방법으로 모형을 검토하였다. 첫째, 비정상/정상 자료여부를 검토하였다. 둘째, 자료의 정상성 여부 또한 자기상관함수를 살펴보았을 때 월별 자료가 주기마다 스파이크가 표현되어 계절차분을 실시하였다. 셋째, 전문가와 3가지 모형을 정성적으로 평가하였다. 위와 같은 배경 하에 최종적으로 계절 ARIMA 모형 3개를 도출하였다.

모형별 정확도 측정결과는 <표 1>과 같다. 3개의 모형 중 ARIMA (0,0,9)(0,1,0)₁₂모형이 정상 R²이 높아 설명력이 가장 높은 것으로 나타났다. Ljung-Box의 유의확률 또한 .310으로 백색잡음으로부터 독립적인 모델로 판단(황경후

외, 2015)되었다. 선행연구에 따르면 모형의 정확도 검증을 할 때 실제 값과 예측치의 오차 값을 파악하기 가장 용이한 MAPE(평균절대백분비오차)값이 가장 낮은 모델은 ARIMA(0,0,9)(0,1,0)₁₂인 것으로 나타났으며, 14.328%로 분석된 MAPE 값은 선행연구 검토 결과에 따라 비교적 정확한 예측이 가능한 모델로 분류 되는 20% 이내의 값을 나타낸다. 선행연구 결과에 따라 MAPE 값이 가장 낮은 ARIMA(0,0,9)(0,1,0)₁₂을 최종 모델로 선정하였다(이충기, 2003).

2) ARIMA 모형 신뢰성 검증

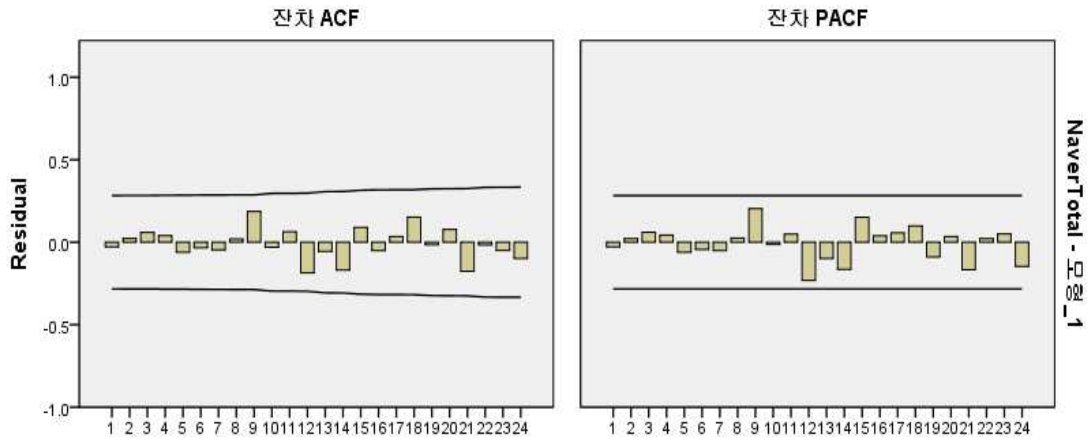
다음 <그림 4>와 같이 잔차(ACF, PACF)의 각 스파이크가 95% 신뢰한계선(신뢰임계구간안)에 포함되어 있다는 것을 알 수 있다. 이는 잔차들의 수치가 대부분 0에 가깝게 나타났기 때문에 계절 ARIMA(0,0,9)(0,1,0)₁₂모형은 통계적으로 유의미(한광중, 2015)한 것으로 판단된다.

3) 모형의 예측

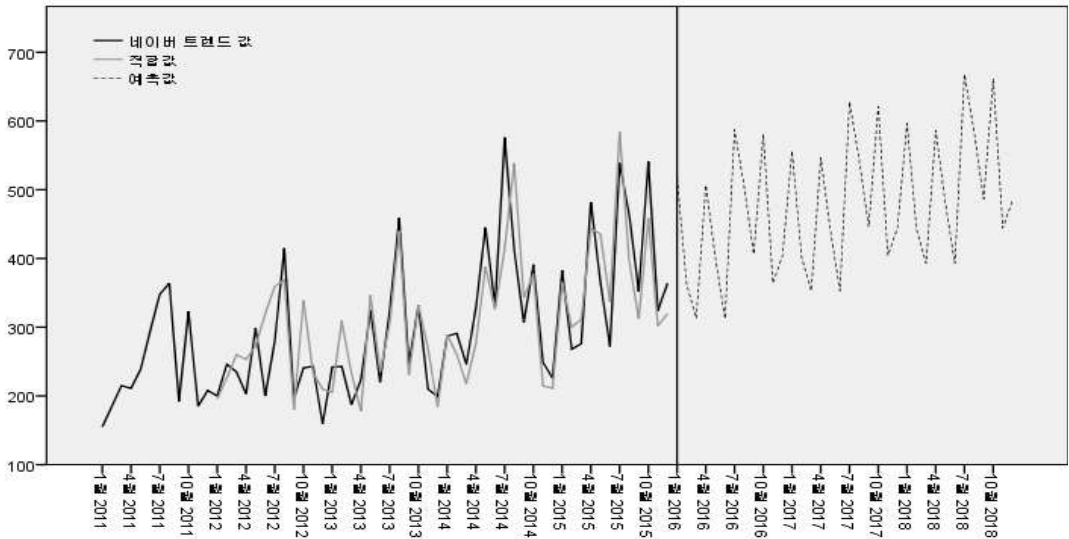
<그림 5>는 텍스트 마이닝 실제 값(2011.01 - 2015.12)과 적합 값 그리고 예측 값(2016.01 - 2018.12)을 그래프로 나타낸 것이다. 적합 값이 예측 값을 유사하게 따라가고 있다는 것을 그림에서 확인할 수 있으며, 이는 MAPE(평균절대백분비오차) 값이 14.328%로 나타난 것으로도 확인할 수 있다. 이러한 분석결과에 따라 안동시 검색 통계 값은 2018년도까지 지속적으로 증가할 것으로 보인다.

<표 1> ARIMA 모형별 정확도 측정 결과

모형	예측 변수	정상 R ²	정규화된 BIC	MAPE	Ljung-Box			이상값수
					통계량	자유도	유의확률	
ARIMA (0,0,3)(0,1,0) ₁₂	0	.095	8.733	15.944	18.047	15	.260	0
ARIMA (0,0,5)(0,1,0) ₁₂	0	.206	8.810	15.895	19.526	13	.108	0
ARIMA (0,0,9)(0,1,0) ₁₂	0	.368	9.004	14.328	10.522	9	.310	0



〈그림 4〉 잔차 비교



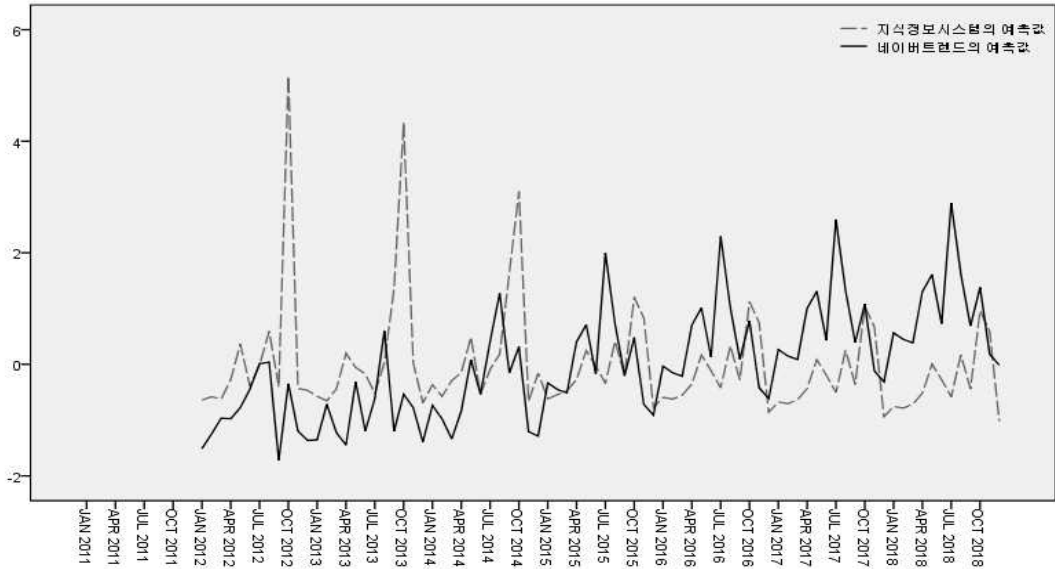
〈그림 5〉 텍스트 마이닝의 실제 값, 적합 값, 예측 값

4) 수요예측 표준화 추세 비교

관광지식정보시스템에서 집계하고 있는 관광객 수는 2013년도부터 관광객 집계지점을 사전 예약제 운영과 입장권 발매, 무인 계측기 설치 등 관광객이 확인 가능한 관광지만 포함하는 형태로 변경되었다. 집계방식 변화는 집계 관광지 감소로 이어져 관광객 수 추이가 2013년을 기점으로 감소하고 있는 것으로 나타났다. 이러한 문제점을 극복하기 위하여 텍스트 마이닝 분석을 통한 수요예측을 실시하기 전, 2011년부터

2014년까지 지속적으로 집계되고 있는 관광지 관광객 수만을 추출하여 수요예측을 하였다. 이는 2013년 이후 집계방식 변화로 관광지 수가 상이하기 때문에 총 관광객 수에 영향을 미칠 것이라 판단하여 지속적으로 관광지 수가 집계되는 관광지만을 수요예측에 활용하였다.

두 데이터를 동일한 기준으로 비교하기 위하여 단위 표준화를 실시하여 순차도표를 출력한 결과는 〈그림 6〉과 같다. 그래프와 같이 관광지 지식정보시스템의 데이터를 사용한 예측 값을 살펴



〈그림 6〉 관광지식정보시스템과 텍스트 마이닝 수요예측 표준화 추세 비교

보면 안동시의 관광객 수는 2018년도까지 지속적으로 감소하는 것으로 예측되었다. 반면, 본 연구의 ARIMA(0,0,9)(0,1,0)₁₂모형을 통한 텍스트 마이닝 수요예측 결과는 2018년도까지 안동시 관광객 수가 지속적으로 증가할 것으로 예측되어 관광지식정보시스템 수요예측과 차이가 있는 것으로 나타났다.

5) 텍스트 마이닝을 통한 수요예측 결과 변환

2016년 1월부터 2018년 12월까지 ARIMA(0,0,9)(0,1,0)₁₂모형으로 텍스트 마이닝 예측 값을 도출한 후, 예측 값의 단위를 〈수식 1〉과 〈수식 2〉를 적용하여 실제 방문객수로 단위 변환을 실시하였다. 수식 수치를 살펴보면 〈수식 1〉에서 SD(표준편차)는 126.53이며 μ(평균)는 382.46을 대입하였다. 〈수식 2〉에서 μ(평균)는 431,637.08, SD(표준편차)는 378,156.35

를 넣어 실제 관광객 수를 도출하였다.

$$Z = \frac{X - \mu}{SD} \quad X = \mu + (Z \times SD)$$

〈수식 1〉 표준화 값 〈수식 2〉 실제값의 변환값

〈표 2〉는 예측치를 연도별로 합산하여 아직 집계되지 않은 2016년부터 2018년까지 방문할 것으로 예상되는 관광객 수를 제시한 값이다. 실제 방문객수로 단위 변환을 실시한 결과 2016년에 안동시 방문객은 약 730만 명, 2017년에는 약 870만 명, 그리고 2018년에는 약 1,000만 명의 관광객이 안동시를 방문할 것으로 예측되었다.

〈표 2〉 연도별 관광객 수 예측 결과

구분	2016년	2017년	2018년
관광객 수	7,321,689명	8,727,135명	10,132,580명

V. 논의 및 시사점

본 연구는 관광개발 및 정책 수립 시 기초자료로 활용되고 있는 수요예측 분석이 국가기관 및 지자체에서 제공하는 통계에 의존하여 현실적인 정보 수집이 어려운 실정이라고 판단하였다. 이에 보다 정확한 수요예측을 실시하기 위해 온라인 검색 엔진을 활용한 텍스트 마이닝을 통해 수요예측을 실시하고자 하였다. 온라인 검색 엔진은 관광객 정보수집 시 가장 많이 사용되고 있는 정보탐색 방법 중 하나로 다양한 정보탐색 행동은 최종적으로는 관광지 선택에 직접적인 영향을 미친다. 따라서 이러한 온라인검색 엔진을 활용한 수요예측이 보다 실질적인 정보를 제공할 수 있을 것이다.

온라인 검색엔진을 활용한 수요예측 결과는 다음과 같다. 첫째, 텍스트 마이닝을 통한 키워드 검색 추이와 관광지식정보시스템 통계자료를 활용한 관광객 수 추이와의 유의성 판단을 위해 두 데이터의 표준화 수치를 비교 분석하였다. 그 결과 두 데이터 간 유사한 추이를 보이고 있으나, 키워드 검색 추이가 높게 나타나는 시기와 관광지식정보시스템 관광객 수 추이가 높게 나타나는 시기가 일치하지 않는 것으로 나타났다. 이는 관광지식정보시스템 통계가 특정 관광지의 관광객 수만을 집계하기 때문에 새로운 관광지 등장으로 새로운 관광지 관광객 수를 반영하지 못하기 때문이라고 판단할 수 있다. 둘째, 텍스트 마이닝 분석을 통해 도출된 키워드 검색 추이를 활용하여 수요예측을 실시한 결과, 안동시 관광객 수가 2018년까지 지속적으로 증가할 것으로 나타났으나, 관광지식정보시스템 통계를 활용한 수요예측 결과에서는 감소할 것으로 예측되었다. 이는 관광지식정보시스템 통계는 2012년까지 유료 및 무료 관광지 통계를 집계하였으나 2013년부터 특정 관광지 관광객만을 집계하기 때문에 전체적으로 관광객 수가 감소하였을 것으로 해석

할 수 있다. 따라서 이러한 관광지식정보시스템 통계 집계 방식 변화가 수요예측에 영향을 미쳤을 것으로 판단된다. 이에 안동시 인기 검색어를 도출하는 텍스트 마이닝 분석은 전반적인 안동시 이슈 키워드를 통한 잠재관광객 수를 추정할 수 있어 보다 유용한 자료로 활용될 수 있을 것이다.

본 연구의 분석결과를 바탕으로 도출된 이론적 시사점은 다음과 같다. 첫째, 본 연구는 텍스트 마이닝을 통해 도출된 키워드 검색 추이와 관광지식정보시스템 관광객 수 추이를 비교 분석한 결과 두 추이가 유의하게 나타나 관광객 의사결정과정에서 관광지를 방문하기 전 온라인 검색 엔진을 통한 정보탐색 행동이 관광지 방문과 직접적인 연관성이 있음을 밝혔다. 이는 현용호 외(2004)의 연구에서 관광 웹사이트 탐색을 통해 정보를 비교 분석한 후 관광지 선택을 결정하게 된다는 연구결과를 뒷받침 하는 결과이다. 둘째, 기존 수요예측은 국가 및 지자체에서 제공하는 통계에 의존하였으나 본 연구에서는 기존방식에서 벗어나 잠재 관광객이 온라인 사이트에서 실제 검색한 검색 추이를 가공한 수요예측을 실시하여 관광 분야 수요예측 분석방법을 확장하였는데 그 의의가 있다. 기존 수요예측은 관광지식정보시스템에서 제공하는 통계자료 및 지자체 집계 방문객 수를 활용한 수요예측 방법이 대부분이었으나 이는 관광객 중복 집계라는 의견이 발생하며 수요예측에 대한 논란이 끊이지 않았다. 그러나 온라인 검색 엔진을 활용한 수요예측은 관광객이 관광지를 방문하기 전 직접 이용한 정보를 기반으로 데이터를 추출하기 때문에 관광객 수를 추정하는데 있어 기존 수요예측 보다 단기 예측에 있어 그 예측력이 더 뛰어날 것으로 판단된다.

본 연구의 실무적 시사점은 다음과 같다. 첫째, 텍스트 마이닝을 통한 데이터 구축은 국가 및 지자체에서 제공하는 통계정보에 의존하는 기존 수요예측 보다 분석시간이 짧기 때문에 특정

사건이 발생하였을 경우 관광시장을 예측하는데 있어 보다 효과적으로 활용될 수 있다. 기존 수요예측은 관광지식정보시스템 통계 및 지자체 통계를 이용하는데 이러한 통계 데이터는 지속적으로 수집되는 것이 아닌 정기적인 업데이트를 통해 자료를 제공하기 때문에 데이터 수집 시 시간이 다소 소요된다. 이에 본 연구는 온라인 검색 엔진에서 최신자료를 짧은 시간에 얻을 수 있는 텍스트 마이닝을 통한 키워드 검색 추이 데이터를 확보하여 수요예측을 실시하였다. 이러한 빅데이터는 온라인 검색 엔진에서 제공하는 정보를 가공하여 사용하여 짧은 시간에 다양한 정보를 얻을 수 있을 것이다. 둘째, 관광지식정보시스템 통계 추이와 네이버 트렌드 검색 수 추이를 비교한 결과 유사한 추이를 보였으나 추이가 높아지는 시기는 다소 차이가 있는 것으로 나타났다. 이는 온라인검색 엔진을 활용한 텍스트 마이닝 검색 추이 데이터가 특정 관광현상을 반영할 수 있으므로 기존 수요예측을 위해 사용되는 통계정보보다 더 우수한 예측력을 지니고 있다고 볼 수 있다. 셋째, 관광지식정보시스템은 특정 관광지 관광객만을 집계하기 때문에 실제 관광객 관광행태를 반영하지 못하는 것으로 판단된다. 관광객은 다양한 관광지를 방문하지만 관광지식정보시스템은 관리 가능한 관광지 관광객만을 집계하기 때문에 실제 지역을 방문한 관광객과 상이한 결과를 나타낸다. 이에 기존 수요예측 방식에서 벗어나 온라인 검색 엔진을 통해 도출된 빅데이터를 활용한 수요예측이 더 실효성이 높을 것이라 판단할 수 있다.

그러나 본 연구는 위와 같은 이론적·실무적 시사점을 가지고 있음에도 불구하고 다음과 같은 한계점을 지니고 있다. 온라인 검색 엔진 키워드를 활용한 시계열 분석을 실시하여 단기 예측은 비교적 예측력이 높으나 장기 예측 시 신뢰도가 떨어질 수 있다는 한계점을 지닌다. 또한 본 연구에서 활용한 검색 엔진은 전체 검색량을 0-

100단위 비율로 변환하여 수치를 제공하고 있기 때문에 실질적인 검색 횟수는 확보하기 어려워 다양한 검색엔진 검색량의 정량적인 취합이 어렵다는 한계점을 지닌다. 따라서 향후 연구에서는 예측오차를 줄일 수 있는 정교한 분석 방법이 시도되기를 기대한다.

참고문헌

관광지식정보시스템(2015). <http://www.tour.go.kr>
 경북일보(2016.01.12). 안동시, 1천만 관광객 시대 연다.
<http://www.kyongbuk.co.kr/?mod=news&act=articleView&idxno=947533>
 교통개발연구원(1988). 『장기관광수요예측에 관한 연구』. 세종.
 김성근·조혁준·강주영(2016). 학술연구에서의 텍스트 마이닝 활용 현황 및 주요분석기법. 『정보기술아키텍처연구』, 13(2), 317-329.
 김휴찬·김미정·신현정(2014). 소셜데이터 분석방법을 이용한 제주관광 트렌드 서비스 설계 및 구현. 『e-비즈니스연구』, 15(3), 173-193.
 노컷뉴스(2016.10.13). 여수 부시장 관광객 1300만, 실제 훨씬 적다.
<http://www.nocutnews.co.kr/news/4668545>
 대구신문(2016.03.21). 안동, 체류형 관광객 늘었다.
<http://www.idaegu.co.kr/news.php?code=tk0301&mode=view&num=192314>
 문화체육관광부(2014). 『빅 데이터와 인터넷 검색 기술을 이용한 속보성 있는 관광수요 추정 기술 개발』. 서울.
 서정호·김철원(2012). 국내관광 의사결정과 소셜 미디어의 관계성 규명: 내용분석을 중심으로. 『관광학연구』, 36(10), 125-148.
 손은호·서진우·정명보(2005). ARIMA 모형을 이용한 호텔식음료 매출액의 예측. 『관광·레저연구』, 17(3), 117-132.
 송근석·이혜숙(2006). 관광수요예측모형의 정확성

- 비교에 관한 연구: 단변량 시계열모형을 중심으로. 『관광정보연구』, 23, 83-111.
- 심진옥·정철(2016). 여행상품 옵션선택에서의 옵션 프레이밍 효과: 정보제시 유형을 중심으로. 『관광학연구』, 40(8), 23-41.
- 심영석·김홍범(2016). 텍스트 마이닝을 이용한 관광지 이미지 구성요인 및 측정에 관한 연구. 『관광학연구』, 40(7), 221-245.
- 안명숙·민용기(2016). 5성급 호텔 이용에 대한 동기 빅데이터 분석: 동기이론 push-pull factor를 중심으로. 『관광연구』, 31(5), 87-86.
- 안창원·황승구(2012). 빅 데이터 기술과 주요 이슈. 『정보과학회지』, 30(6), 10-17.
- 이시우(2002). 국외 관광객 관광상품 정보탐색에 관한 연구. 『관광경영연구』, 14, 190-215.
- 이충기(2003). 『관광응용경제학』. 서울: 일신사.
- 전북도민일보(2013). 전북 관광객 뺑뚱기 홍보, 내년부터 안 통한다.
<http://www.domin.co.kr/news/articleView.html?idxno=1007233>
- 전창석(2003). 외부정보형태별 정보 탐색량과 여행사선택행동에 관한 연구. 『관광·레저연구』, 15(1), 79-95.
- 정병욱(2013). 관광블로그 품질이 사용의도에 미치는 영향: 확장 기술수용모델 및 IS 성공모형을 중심으로. 『관광연구논총』, 25(2), 81-109.
- 정철(2009). 관광정보 노출 효과. 『관광연구논총』, 21(2), 39-60.
- 정철·이준남(2010). 인터넷 관광정보에 대한 사용자 지각, 태도, 그리고 정보탐색노력. 『관광학연구』, 34(5), 265-286.
- 조완섭·조아·권가은·류관희(2015). SNS 데이터 분석을 통한 스마트 충북관광 구축. 『한국데이터정보과학회지』, 26(2), 409-418.
- 중부일보(2014.04.21). 파주시 관광객 1천만 명 시대, 뺑뚱기 집계 논란.
<http://www.joongboo.com/news/articleView.html?idxno=918462>
- 최병길·오상훈·현정석(2006). 여행형태별 제주관광수요 패턴 및 예측. 『관광·레저연구』, 18(4), 41-61.
- 표원정·임관혁(2012). SNS 관광정보의 특성이 관광객 정보탐색행동에 미치는 영향. 『관광경영연구』, 16(2), 285-308.
- 한겨레신문(2015.08.21). 지역경제 살리는 열차 난민 '내일러'를 아십니까.
http://www.hani.co.kr/arti/economy/economy_general/705468.html
- 한광중(2015). 『SPSS 활용 미래예측과 시계열 분석』. 서울: 백산출판사.
- 한진성·윤지환(2016). 소셜 빅데이터 텍스트 마이닝을 활용한 부산국제영화제(BIFF) 활성화 방안 연구: 영화제 개최 전·후 비교 분석을 중심으로. 『관광학연구』, 40(1), 133-145.
- 허향진·송재호(2000). 시계열을 활용한 제주지역 관광객 수요 예측 예측모델 간 비교와 유치목표치 설정. 『제주관광학연구』, 3, 93-105.
- 현용호·허향진·김윤숙(2004). 온라인 관광객의 사결정과정에서 웹기능에 관한 탐색적 연구. 『관광학연구』, 28(2), 291-312.
- 황경후·김원표·정철(2015). 일변량 시계열 모형 비교를 통한 독도관광 수요예측. 『관광·레저연구』, 27(2), 59-77.
- 황현철·박재현(2000). 관광정보 탐색 행동유형에 관한 연구. 『산업경제연구』, 13(6), 1-18.
- Artola, C., Pinto, F., & de Pedraza Garcia, P. (2015). Can internet searches forecast tourism inflows? *International Journal of Manpower*, 36(1), 103-116.
- Box, G., & Jenkins, G. (1976). *Time series analysis: Forecasting and control*. San Francisco: Holden-Day.
- Burger, C., Dohnal, M., Kathrada, M., & Law, R. (2001). A practitioners guide to time-series methods for tourism demand forecasting: A case study of Durban, South Africa. *Tourism Management*, 22(4), 403-409.
- Fodness, D., & Murray, B. (1999). A model of tourist information search behavior. *Journal of Travel Research*, 37(3),

- 220-230.
- Ginsberg, J., Mohebbi, M., Patel, R., Brammer, L., Smolinski, M., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature*, 457(7232), 1012-1014.
- Gursoy, D., & McCleary, K. (2004). An integrative model of tourists' information search behavior. *Annals of Tourism Research*, 31(2), 353-373.
- Heerschap, N., Ortega, S., Priem, A., & Offermans, M. (2014). Innovation of tourism statistics through the use of new big data sources. In *12th Global Forum on Tourism Statistics*, Prague, CZ.
- Kitchin, R. (2014). The real-time city? Big data and smart urbanism. *GeoJournal*, 79(1), 1-14.
- Labrinidis, A., & Jagadish, H. (2012). Challenges and opportunities with big data. *Proceedings of the VLDB Endowment*, 5(12), 2032-2033.
- Lau, K., Lee, K., & Ho, Y. (2005). Text mining for the hotel industry. *Cornell Hotel and Restaurant Administration Quarterly*, 46(3), 344-362.
- Lee, C., Song, H., & Mjelde, J. (2008). The forecasting of international expo tourism using quantitative and qualitative techniques. *Tourism Management*, 29(6), 1084-1098.
- Lim, C., & McAleer, M. (2002). Time series forecasts of international travel demand for Australia. *Tourism Management*, 23(4), 389-396.
- Loh, S., Lorenzi, F., Saldaña, R., & Lichnow, D. (2003). A tourism recommender system based on collaboration and text analysis. *Information Technology & Tourism*, 6(3), 157-165.
- Ma, E., Liu, Y., Li, J., & Chen, S. (2016). Anticipating Chinese tourists arrivals in Australia: A time series analysis. *Tourism Management Perspectives*, 17, 50-58.
- Mathieson, A., & Wall, G. (1982). *Tourism, economic, physical and social impacts*. London: Longman.
- Preis, T., Moat, H., & Stanley, H. (2013). Quantifying trading behavior in financial markets using Google trends. *Scientific Reports*, 3.
- Taylor, D., Lewin, J., & Strutton, D. (2011). Friends, fans, and followers: Do ads work on social networks? *Journal of Advertising Research*, 51(1), 258-275.
- Wang, Y., & Fesenmaier, D. (2006). Identifying the success factors of web-based marketing strategy: An investigation of convention and visitors bureaus in the United States. *Journal of Travel Research*, 44(3), 239-249.
- Xiang, Z., & Gretzel, U. (2010). Role of social media in online travel information search. *Tourism Management*, 31(2), 179-188.

2016년 9월 20일 최초투고논문 접수

2016년 10월 20일 최종심사완료 및 게재확정 통보

2016년 11월 10일 최종논문 도착

3인 익명심사 畢