

여행자 관심 기반 스마트 여행 수요 예측 모형 개발: 웹검색 트래픽 정보를 중심으로*

박도형**

<목 차>

I. Introduction	III. Research Design and Results
II. Theoretical Backgrounds	3.1 Qualitative Research on Information Search Behavior of Potential Travelers
2.1 Importance of Travel Demand Prediction	3.2 Travel Demand Model Based on Web Search Traffic Data
2.2 Web Search Traffic Information	IV. Conclusion
2.3 Travel Demand Nowcasting Based on Web Search Traffic Data	참고문헌
	<Abstract>

I. Introduction

정보 기술의 발전과 이에 상응하는 데이터 수집/분석 기법의 발달에 따라 빅데이터라고 불리는 화두가 최근 주목받고 있다. 빅데이터 분석은 데이터 수집, 저장, 가공 등의 선행 활동뿐만 아니라 데이터 추출, 변수 선정, 모형 개발의 중심활동까지 다양하게 발전하고 있다. 각자 대상은 다르지만, 공통적으로 기존에 발견하지 못한 새로운 가치 창출을 목적으로 하고 있다(강태영 & 박도형, 2016). 과거 정보시스템 자원의

비용이나 한계 등으로 활용성이 분명치 않으면 수집되지 않았던 정보들도 최근에는 수집/보관되고 있으며, 오픈소스나 오픈데이터 등의 개방형 혁신 트렌드에 따라 연구자나 일반인에게 공개되고 있다. 기업이나 기관 내부에서 개인 정보보호 아래 자체적으로 관리되고 있는 소비자 개인 정보, 구매 및 판매 기록 정보, 물류나 유통 정보 및 제품이나 서비스 로그 정보 등은 접근이 제한되나, 온라인 상에 공개되어 존재하는 뉴스 정보, 사용자 피드백 정보, 웹검색 트래픽 정보 등은 연구자의 기호에 맞게 수집될 수

* 이 논문은 '2015년 한국은행 강원본부 공동연구과제'로 수행된 연구를 기반으로 기관의 승인 아래 작성된 논문으로서, 과제에는 저자 외에 이민섭 조사역이 참여했음.

** 국민대학교 경영대학/비즈니스IT전문대학원 부교수, dohyungpark@kookmin.ac.kr(주저자, 교신저자)

있다. 이 중에서 포털이 제공하고 있는 웹검색 트래픽 정보는 일반인들의 관심사항이나 관심의 변화, 사전 수요 등을 반영하고 있으며, 실시간까지는 아니지만, 비교적 최근의 정보가 지속적으로 제공되기 때문에 활용성이 매우 크다고 할 수 있다(전승표 & 박도형, 2013; 김범수 외, 2015).

제품이나 서비스를 소비하는 고객들 대부분은 구매의사결정단계를 따르며 본인의 소비를 최적화 하고자 하며, 이 단계에서 정보탐색은 소비 대상을 선정하는 데 매우 중요한 단계이다(이동원 외, 2014). 정보탐색 활동 중 많은 부분은 인터넷 상에서 정보를 검색하는 것이고, 일반 개인들은 인터넷 검색 활동이 효율적이고 합리적인 구매 결정에 필수불가결한 단계로 인식하고 있다(Jun et al., 2014). 인터넷 정보검색 활동은 제품이나 서비스 구매시에 개인이 갖게 되는 불확실성을 감소시킬 수 있는 가장 비용 효율적인 행위이고, 검색 방법이나 기법 등의 검색 노하우는 합리적 소비를 위한 중요한 역량으로 인식되고 있다. 기업이나 기관 또한 소비자가 검색활동을 하는 동안 자신의 제품이나 서비스가 노출되는 것이 매우 중요하기 때문에, 많은 비용을 검색 마케팅에 지출하고, 여러 방법을 통해 검색시 상위에 노출되도록 전략을 펼치고 있다. 이들은 소비자가 검색시 자신이 전달하고자 하는 내용을 광고나 홈페이지를 통해 노출시키기도 하고, 이보다 더 효과적이라고 알려진 사용자 후기 형태의 온라인 구전을 노출시켜 소비자에게 자사의 제품이나 서비스를 어필하고 있다(Park et al., 2007; Park & Kim, 2008; Park & Lee, 2008). 인터넷 검색이 기업이나 기관의 설득수단으로 보여질 수도 있으나,

소비자 입장에서는 최종적으로 확실하고 정확한 정보가 제공되어 제품이나 서비스에 대한 만족이 증가할 수 있는 바람직한 수단의 역할을 하고 있다(전승표 & 박도형, 2013). 구글을 시작으로 인터넷 포털 서비스 기업들은 점차적으로 자신이 축적하고 있는 웹검색 트래픽 정보를 제공하고 있다. 정확한 검색량은 키워드 광고를 신청할 시에만 확인할 수 있지만, 상대적인 증감에 관해서는 온라인 상에서 간단히 확인할 수 있도록 하고 있다. 웹검색 트래픽 정보가 소비자의 관심을 확인하는 데 효과적이며, 어떤 대상에 관심을 가지는지, 관심의 정도는 시간에 따라 어떻게 달라지는지 등의 고급 정보도 제공할 수 있기 때문에 학계 업계에서 많이 활용되고 있다(Jun & Park, 2017). 독감 등의 전염병 예측(Ginsberg et al., 2009), 부동산 시세 변화(Choi & Varian, 2012), 금융 관련 시장의 움직임(김범수 외, 2015), 제품의 수요(Jun & Park, 2016) 등 다양한 방면에서 웹검색 트래픽을 활용한 연구들이 등장하고 있으나, 현재까지 여행수요를 예측하는 데 웹검색 트래픽을 활용한 연구는 부족한 실정이다.

여행자가 여행을 준비할 때, 여러 단계를 거치게 된다. 우선 잠재 여행자들은 여러 다양한 경로를 통해 특정 지역이나 도시에 대한 정보에 대해 노출된다. 주변 지인이 추천하거나, TV 프로그램에서 매력적으로 소개되거나, 이벤트나 축제를 통해 유명세를 갖고 있는 등의 다양한 이유로 후보 여행지에 관심을 가지게 된다. 본인이 가지고 있는 기준을 기반으로 과거 여행 경험, 타인의 의견 등을 종합하여 후보 여행지들을 평가하게 된다. 평가과정 후, 특정 후보 여행지가 자신이 원하는 기준에 부합할 것이라

는 결론에 도달하면 여행갈 준비를 하게 될 것이다. 준비 과정에는 여러 정보가 필요한데, 대부분은 인터넷 검색을 통해 정보를 조달할 것이다. 구체적으로 생각해 보면, 여행자들은 인터넷을 통하여 가고자 하는 지역을 중심 키워드로 여러 다양한 키워드를 만들고 이를 검색한다. 예를 들어, 로마 여행을 고려시, 여행자는 ‘로마 날씨,’ ‘로마 맛집,’ ‘로마 볼거리’ 등의 다양한 키워드를 생성하고 정보를 수집하여 구체적인 계획을 수립하게 된다. 즉, 특정 지역의 검색량이 많다는 것은 해당 지역을 방문할 여행자들이 많아서 잦은 검색이 수행되었다고 유추할 수 있고, 이는 여행자의 수요를 예측하는데 충분히 활용될 수 있음을 시사하고 있다.

본 연구에서는 대한민국 검색서비스 1위 업체인 네이버의 웹검색 트래픽 데이터를 기반으로 강원도 대표 관광도시를 대상으로 여행수요를 예측하는 모형을 개발하고자 한다. 여행수요 예측의 정확도는 정책 수립 및 관련 여행 업체들의 수익에도 중요한 변수로서, 정확한 예측치는 경영의사결정의 기초를 제공하는 중요한 정보이다. 여행수요 예측의 중요성이 강조되고 있으나, 현실에서 활용되는 여행수요 추정은 비교적 최근의 데이터가 반영되지 못하고 있고, 방법 측면에서도 객관적이고 체계적이지 못하다는 지적을 받고 있다. 본 연구에서 제안하는 웹 검색 트래픽 정보는 최신의(거의 실시간에 가까운) 데이터를 활용하며, 여행자의 정보 탐색 그 자체를 반영하기 때문에, 좋은 대안이 될 수 있을 것으로 생각된다. 또한, 대규모 서버이나 관련 업체로부터 데이터를 제공받지 않아도 되기 때문에, 데이터 수집 비용이 무척 적으며, 예측 결과의 정확도를 확보한다면, 경제적으로 효

과적인 방법이라 할 수 있다. 본 연구는 이론적으로 여행수요 예측에 새로운 변수를 사용한다는 점, 실무적으로 강원도 해당 도시의 업체나 기관이 활용할 수 있는 실질적인 방법을 제안한다는 점에서 의의를 가지며, 정책적으로도 정확한 여행수요 예측치 기반의 과학적인 정책 수립에 기여할 수 있다.

II. Theoretical Backgrounds

2.1 Importance of Travel Demand Prediction

우리나라를 비롯한 많은 국가들이 관광산업을 육성하고 있는데, 관광산업의 성장이 GDP와 고용률에 미치는 직/간접효과가 분명하며 이와 더불어 내수 증진이나 외자 유치 등의 경제적 파급효과도 기대 이상이기 때문이다(송근석 & 송학준, 2006). 관광산업의 지속적 발전과 경쟁력을 확보하기 위해서는 정확한 데이터 기반의 과학적이고 합리적인 의사결정이 필요하다. 여기에서 의사결정의 기본이 되는 일반 개인들의 여행수요 예측치는 매우 중요한 역할을 수행한다. 예를 들면, 여행수요를 적게 추정할 경우, 항공좌석 등의 교통시설 부족이나 호텔객실이나 노동력 등의 서비스 부족으로 수익을 창출할 수 있는 기회를 상실하게 되고, 여러 시설이나 설비에 대한 수용력이 초과되어 여행자들은 만족할만한 서비스를 받지 못하고 불만을 가지고 떠나게 될 것이다. 반면에 과다하게 예측할 경우, 교통, 숙박, 서비스 면에서 과도하게 비용이 집행되어 손실이 발생할 것이다. 이처럼

수요예측의 정확성은 공공과 민간부문의 의사 결정에 막대한 영향력을 가지기 때문에, 관광 관련 연구에서 상당수가 여행수요 예측모형을 개발하고 그것의 필요성과 정확성을 강조하고 있다(Sheldon & Var, 1985; Uysal & Crompton, 1985; Archer et al., 1987; Dharmaratne, 1995; Witt & Witt, 1995; Song et al., 2003). 이들 연구에서는 여행이 서비스인 만큼 서비스의 대표 특징인 소멸성 관점에서 정확한 수요예측이 중요하다 하였고, Song et al.(2003)은 여행수요 예측은 중앙정부와 지방 정부의 공무에 중요한 업무이고, 정확성과 실제 사용성이 높은 수요예측 모형을 개발할 필요를 역설하였다.

강원도는 사회 인프라 구축 및 국내외 여행 수요 증가로 인해 대한민국 대표 관광지로 자리매김하고 있다. 특히 2018년 평창 동계올림픽을 고려하면 강원도 여행수요는 현재보다 더 크게 증가할 것으로 판단된다. 강원도가 여행자에게 양질의 서비스 제공을 위해서, 정확한 여행수요 예측에 기반한 유형/무형 자산개발 계획과 실행이 당장 필요해 보인다. 또한, 장기적인 관점에서 대한민국 대표 관광지의 명맥을 이어가기 위해서, 시설 구축, 문화 개발, 인력 고용 등에 예산을 투입하고 집행하는 것이 필요한데, 정확한 여행수요 예측 자료는 중요하게 사용될 수 있다. 또한 향후 예정된 평창 동계올림픽 개최 이전의 각 분야별 시설 및 인력 예산 집행과 개최 후 시설 관리, 이용 전략 수립 및 운영을 위해서도 관련 자료는 중요하게 사용될 수 있다. 인천공항철도 이용 실적이 개통 당시 수요 예측에 비해 턱없이 부족하다는 점, 인천 아시안게임을 위한 신규 설립된 경기장 및 시설의

수요 부족과 과다한 비용 발생한 점 등은 정확한 수요예측이 없는 상태에서 예산이 집행될 경우의 피해를 여실히 보여주고 있다.

2.2 Web Search Traffic Information

최근 인터넷 검색 서비스 업체가 검색 통계를 무료로 오픈하면서, 이를 활용한 연구가 다양하게 진행되고 있다. 웹검색 트래픽 정보의 예측 가능성이 본격적으로 알려지기 시작한 첫 연구는 Ginsberg et al.(2009)의 독감 바이러스의 징후를 예측한 논문이다. 이 논문에서는 구글이 2008년 공개한 서비스인 'Flu Trends'로부터 얻은 데이터의 분석 결과가 제시되어 있는데, 일반 개인들이 감기에 걸릴 때 검색하는 여러 징후들의 키워드를 바탕으로 독감 바이러스에 대한 염려를 예측하는 서비스이다. 인상적인 점은 미국 내 질병통제센터의 공식 발표보다 1~2주 먼저 독감 바이러스의 전파를 정확히 예측할 수 있다는 점이다. Butler(2013)는 Ginsberg et al.(2009)의 연구를 비판하고 있지만, Ginsberg et al.(2009)이 웹검색 트래픽을 활용해서 긴급하고 중요한 사회 현상을 빠르게 감지할 수 있는 가능성을 보여 준 점, 쉽고 편리하게 모니터링 할 수 있는 점을 보여주었다는 측면에서 의미가 크다고 할 수 있다.

Choi & Varian(2012)의 연구는 단기간의 경제적 예측에 대해 검색 트래픽을 활용해 접근한 것으로 중요한 시사점을 가지고 있다. 구글 트렌드는 사람들의 현재 경제 활동을 예측할 수 있는 것으로 나타났는데, 이 연구에서는 경제활동을 자동차 판매, 주택 판매, 및 소매 등의 카테고리로 설명하고 있다. 여러 카테고리에서

예측력을 시험해 본 결과, 웹검색 트래픽 정보를 활용한 예측은 먼 미래보단 현재에 가까운 미래를 예측하는 데에 유용하다고 보고하고 있다. 이와 더불어 Lui et al.(2011)에서는 2008, 2010년 미국의 의회 선거에서 구글 트렌드의 검색 트래픽을 활용해 당선자를 예측할 수 있는 가능성 주제로 연구한 논문을 발표했다. Vosen & Schmidt(2011)의 연구는 미시건 대학이 발표하는 소비자 심리지수와 신뢰지수를 웹검색 트래픽 총계와 비교하여 웹검색 트래픽 정보의 유용성을 주장하고 있다.

2.3 Travel Demand Nowcasting Based on Web Search Traffic Data

관광산업에서 여행수요 예측은 여행에 종사하고 있는 지역 주민뿐만 아니라 지역 여행 정책을 수립하는 여러 정부 기관에도 중요하다. 이유는 여행수요 예측 결과를 기반으로 관련 정책이나 해당 산업의 사업 계획을 수립할 수 있기 때문이다. 이에 따라 여행 관련 많은 연구들은 여행수요를 정확히 예측하기 위한 방법론이나 예측모형을 꾸준히 제공하고 있다.

여행수요 예측 방법은 크게 정량적 방법과 정성적 방법으로 분류되고 있으며 이 둘을 통합하는 결합방법이 포함되기도 한다(Uysal &

Crompton, 1985; Witt & Witt, 1995; Song et al., 2003). <표 1>은 두 방법의 차이를 제시하고 있다.

여행수요를 예측하는 많은 연구들은 계량적 기법을 사용하여 수요함수를 추정하고 있다. Sims(1980)의 연구에서 제안된 Vector Autoregressive 모형과 Vector Error Correction 모형이 시계열 분석기법으로 주로 활용되고 있고, Oh(2005)는 분기별 자료를 가지고 Vector Autoregressive 모형을 개발해 국내 여행수요를 설명했으며, Song & Witt(2006)는 마카오 여행수요 함수를 동일한 기법으로 추정하였는데, 마카오 공항 개장, 사스(SARS), 금융위기를 변수로 포함시켰다. Song et al. (2003)의 연구는 여행수요의 중요한 결정 요인으로 소득을 뽑았으며, 이 외에도 물가, 환율 등도 중요한 요인으로 밝혀냈다. 이들 연구를 종합하면, 여행수요 관련 연구는 (1) 예측력을 높이기 위해 발전된 계량기법을 적용하는 방향, (2) 여행수요를 잘 설명할 수 있는 요인을 찾아 이를 모형에 반영하는 방향으로 발전하고 있다는 것을 알 수 있다.

여행자의 여행지 선정 과정은 Crompton (1979)와 Um(1991)의 연구 등에서 주장하는 순차적인 단계로 이루어진다. 여행객은 일상적인 생활을 하면서, 영화, TV 프로그램 등의 외

<표 1> 정성적 vs. 정량적 여행수요 예측 방법 차이

	정성적 방법	정량적 방법
목적	특정 행동에 대한 원인이나 동기의 질적인 이해를 목표로 탐색적 접근	특정 행동의 데이터를 기반으로 패턴을 발견하는 것을 목표로 확증적 접근
표본	주제 관련 많은 경험 있는 소수	일반 대상 다수
자료수집/분석	비구조화/ 비통계적	구조화/ 통계적
조사방법	심층인터뷰, 델파이기법	설문조사, 다변량 회귀모형, 시계열모형

부 자극이나 사회 문화적 학습, 축적된 과거의 경험 등을 통해 Awareness Set을 형성하고 이를 의식단계라고 한다. 여기서 예산이나 시간 등의 현실적 제약을 반영하고, 정보 탐색 활동을 통해서 가능한 대안을 추려내는데(Evoked Set), 이를 탐색단계라고 한다. 본인이 가지고 있는 기준에 따라, 가능한 대안들을 개별적으로 평가하고 최종 하나의 목적지를 선택하는 단계를 마지막으로 가진다(나운중 & 박은희, 2001).

웹검색 트래픽 정보는 소비자가 적극적인 탐색활동을 하는 결과로서 존재하는 데이터로서, 앞선 여행지 선택 단계에서 탐색단계의 소비자 데이터이다. 따라서 웹검색 트래픽과 여행수요 간에는 정의 상관관계가 있을 것으로 예측할 수 있다. 여행지 선정 과정의 3단계 모형에 따라 정보 탐색 활동은 최종 방문 전에 일어나기 때문에, 웹검색 트래픽 정보는 여행수요보다 선행할 것으로 예상되며, 숙소나 볼거리 등의 여행 계획에 중요한 정보는 먹거리 등의 정보보다 더 먼저 선행할 것으로 생각된다. 여행 동기 문헌들이 주장하는 일반적인 동기(ex. 일상 탈출, 힐링 등)와 이벤트/ 축제 참가 및 생태 체험, 문화 탐방 등의 목적지향 여행 동기 등을 고려하여(강미희 외, 2002), 다양한 키워드의 웹검색 트래픽 정보를 수집할 수 있을 것이고, 이 중 여행수요 예측에 보다 효과적인 키워드가 있을 것이다. 이는 곧 해당 지역의 특징을 반영한 키워드라 생각될 수 있고, 키워드별 웹검색 트래픽 정보는 여행지의 특징을 확인하는 데 활용될 수도 있을 것이다.

III. Research Design and Results

3.1 Qualitative Research on Information Search Behavior of Potential Travelers

본 연구는 정성적인 조사와 정량적인 모형을 함께 수행하여 강원도의 특정 도시의 여행수요를 예측해 보고자 한다. 즉, 여행자들이 여행 계획을 수립할 때, 어떤 목적을 가지고, 어떤 키워드를 통해 온라인 상에서 정보 검색을 하는지를 우선적으로 확인해야 한다. 정보 검색 행태 파악을 위해 본 연구에서는 정성조사를 기획하여 여행 인사이트를 도출하였고, 이를 기반으로 데이터를 수집하여 최종 모형을 도출하였다.

정성조사는 여덟 명의 선발된 대학생들이 자신의 주변 사람 중 여행을 자주 가는 2명 이상을 선정하여 심층 인터뷰를 수행하는 방식으로 진행하였다. 인터뷰는 7일간 실시되었다. 정성조사에서 확인해 보고자 한 내용은 조사 대상자의 여행에 대한 태도 및 목적을 묻는 3문항, 여행 계획 수립시 사전 정보 탐색 활동을 확인하는 6문항, 강원도 여행 관련 3문항으로 구성되어 있다. 총 20여명의 인터뷰 대상자들의 응답을 종합해 보면, 1년에 평균적으로 3회 정도 여행을 다니며, 여행이란 ‘힐링’ 혹은 ‘재충전’의 의미가 있다고 여기고 있었다. 대부분 친구나 가족과 함께 여행을 가는 경우가 많았으며, 여행에 관한 정보는 인터넷을 통해서 검색을 통해 수집하며, 보통 국내의 경우 한 달 정도를 남겨놓은 시점에서 숙소, 볼거리, 축제, 먹거리 등을 탐색한다고 하였다. 많은 경우 대한민국 1

위 검색 업체 ‘네이버’를 가장 많이 활용하였고, 검색 키워드는 가고자 하는 도시에 찾고자 하는 주제를 결합한 조합을 검색키워드로 설정한다고 하였다. 최종적으로 강원도하면 떠오르는 것과 대표 도시를 묻는 질문에서 자연 경관, 군부대, 지역 축제 등이 강원도 하면 생각나는 것이라고 하였고, 평창, 춘천, 속초, 강릉의 순서로 강원도 대표 도시를 뽑았다.

정성조사를 수행한 후 결과를 종합해 보면 다음의 사항이 정량 모형을 개발하는데, 반영되었다.

- 강원도의 대표 도시: 평창
- 검색키워드의 후보군: 평창 + ‘먹거리’, ‘숙소’, ‘볼거리’, ‘축제’
- 여행수요와 웹검색 트래픽 정보의 시차: 3~5주

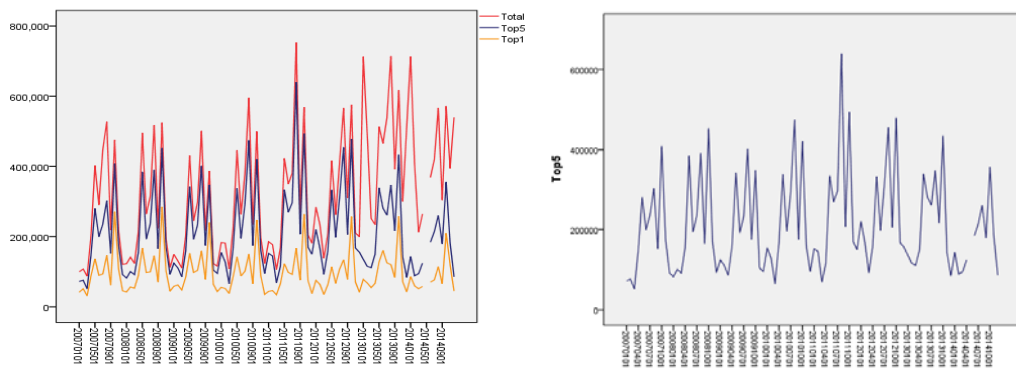
3.2 Travel Demand Model Based on Web Search Traffic Data

웹검색 트래픽 여행수요 예측모형 개발을 위해, 종속변수와 독립변수를 정의하고 해당되는 데이터를 수집하였다. 즉, 강원도의 대표 여행

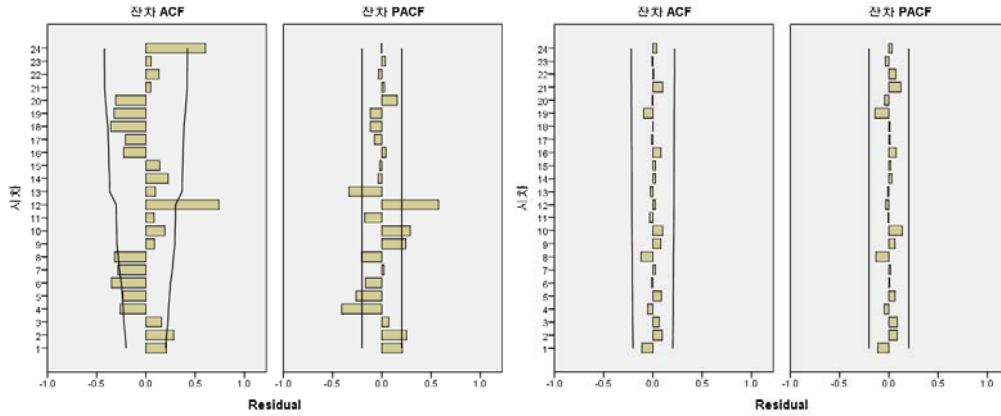
도시로 선정한 2018 동계올림픽 개최 예정지 평창의 여행수요 예측을 위해 평창 지역의 여행수요 데이터와 평창과 관련된 웹검색 트래픽 데이터를 수집하였다.

우선, 종속변수인 평창 여행수요 데이터는 관광지식정보시스템에서 입수 가능한 2007년 1월부터 2014년 12월까지의 여행수요 데이터를 수집하였다. 수집된 데이터는 세 가지 형태로 가공하였는데, 평창 지역 전체 관광명소 입장객을 모두 더한 여행수요를 ‘Total’이라 명명하고, 입장객이 가장 많은 관광 명소의 입장객수를 ‘Top 1’이라 하였고, 입장객이 많은 순서로 다섯 개의 명소의 입장객을 합친 ‘Top 5’이라 하였다. 아래의 <그림 1>은 세 가지 가공된 여행수요의 월별 변화량을 보여주고 있다. 이 중에서 ‘Top 5’ 여행수요를 최종 종속변수로 선정하였는데, 이유는 ‘Top 5’ 여행수요가 다른 변수 대비 상대적으로 정상 시계열에 가깝다고 판단하였기 때문이다.

평창 여행수요는 1년을 주기로 일정한 패턴을 가지고 있으므로 계절성이 존재한다고 할 수 있다. 자세히 보면 연 중 4월에서 9월까지 꾸준히 여행수요가 증가함을 확인할 수 있다.



<그림 1> 평창 관광객 수와 종속변수로 선정된 ‘Top 5’ 여행수요



<그림 2> 여행수요의 자기상관 및 편자기 상관(좌: ARIMA(0,0,0)(0,0,0), 우: ARIMA(1,0,0)(1,0,0))

평창 여행수요 데이터를 정상 시계열로 변환하기 위하여, 자기상관과 편자기상관을 확인하였다(<그림 2> 참조). 열두번째 시차에서 자기상관과 편자기상관 줄다 신뢰한계를 넘고 있고, 앞서 예상한 것처럼 계절차분을 수행해야 한다. 계절차분을 기본으로 하고 여러 시행착오를 거쳐 ARIMA(1,0,0)(1,0,0) 모형이 평창 여행수요 데이터만을 가지고 개발할 수 있는 최적의 시계열 모형이라 판단하였고, 이를 벤치마킹 모형으로 두어, 웹검색 트래픽 데이터가 추가될 때 더 나은 설명력을 가져야 하는 기준점으로 설정하였다.

독립변수인 평창 여행정보 검색 트래픽을 수집하기 위해, 2007년 1월부터 2014년 12월까지 같은 기간의 ‘먹거리’, ‘숙소’, ‘볼거리’, ‘이벤트(명소)’ 관련 키워드 다수의 웹검색 트래픽 데이터를 수집하였다. 해당 데이터는 네이버 트렌드에서 주별 데이터를 추출하고, 월별로 변환하는 과정을 거쳤다. 키워드 생성 방식은 ‘도시명 + 주제’의 형식을 취하였다. 즉, ‘속초 먹거리’, ‘강릉 숙소’ 등의 키워드를 선정한 것이다. ‘강

릉 숙소’의 경우 ‘강릉’은 경포대, 오죽헌 등을 포함한 상위 개념이고 ‘숙소’의 경우도 호텔, 콘도, 민박 등을 포함한 상위 개념이다. 실제로 사람들은 이보다 더 구체적인 수준에서 검색을 하기도 하지만, 이들의 검색량은 주로 상위수준 검색량과 비례관계를 갖는다. 더불어 하위 수준의 단어들을 모두 조합할 경우, 너무 많은 키워드가 도출되어서 연구를 수행하는 데 어려움이 따를 수 있다. 따라서 본 연구에서는 ‘도시명 + 주제’ 레벨에서 키워드를 선정한 것이고, 만약 특정 도시에 대해서 좀더 구체적인 분석이 필요할 경우 수준을 한 단계 낮추어 좀더 다양한 키워드의 조합으로 본 연구의 방법대로 연구를 수행해 볼 수 있을 것이다.

앞서 기술한 정성조사에서 여러 다양한 키워드 후보군을 도출하였는데, 그 중에서 검색량이 충분한지, 일반적인 여행정보 탐색에서 자주 다루어지는지, 특별한 이슈에 의해 검색량의 비정상적인 패턴이 있는지 등을 고려하여 키워드를 선정하였다. 최종으로 도출된 키워드는 ‘평창 숙소’, ‘평창 맛집’, ‘평창 볼거리’ 이다. 선정하

<표 2> 평창 여행수요와 검색키워드 간의 교차 상관.

시차	평창 숙소		평창 맛집		평창 볼거리	
	교차상관	표준오차a	교차상관	표준오차a	교차상관	표준오차a
-7	-.234	.107	-.249	.107	-0.250	0.107
-6	-.167	.106	-.239	.106	-0.154	0.106
-5	-.181	.105	-.217	.105	-0.189	0.105
-4	-.126	.105	-.174	.105	-0.086	0.105
-3	.130	.104	.241	.104	0.322	0.104
-2	-.014	.104	.254	.104	0.171	0.104
-1	.229	.103	.376	.103	0.397	0.103
0	.022	.103	.450	.103	0.364	0.103

<표 3> 여행수요 모형 결과

	Benchmark Model	Web Search Traffic Model
R2	.705	.723
RMSE	69048.244	68004.808
MAPE	24.167	24.236
MaxAPE	155.113	145.450
MAE	49114.590	48414.360
MaxAE	246085.286	267444.301
Standardized BIC	22.429	22.542
Ljung-Box Q	8.209(p = 0.942)	10.92(p= 0.814)
Constant	177958.207 (t = 3.548, p<0.05)	177958.207, (t = 3.548, p<0.05)
AR	0.451(t = 4.805, p<0.05)	0.451(t = 4.805, p<0.05)
Seasonal AR	0.855(t = 17.994, p<0.05)	0.855(t = 17.994, p<0.05)
Traffic for Sightseeing	-	397.167(t = 2.338, p<0.05)
Traffic for Food	-	-96.357(t = -0.481, p>0.05)
Traffic for Accommodation	-	-182.572(t = -1.195, p>0.05)

키워드의 검색 트래픽이 실제 여행수요 대비하여 어느 정도 선행하는지 확인하기 위해 교차 상관 분석을 수행하였다 (<표 2 >참조).

‘볼거리,’ ‘맛집,’ ‘숙소’ 중 ‘맛집’을 제외하고는 한 달의 시차를 가지고 여행수요와 검색 트래픽 간의 관계가 가장 높다는 것을 알 수 있었다. ‘맛집’의 경우 교차 상관 결과에서는 당

월 데이터가 가장 적합하다고 나왔으나, 예측 모형에는 전기의 데이터를 활용하는 것이 더 좋다고 판단되어 차선책인 한 달의 시차를 둔 웹검색 트래픽 데이터를 독립변수로 사용하기로 하였다.

최종적으로 웹검색 트래픽 정보가 반영된 예측 모형의 설명력을 확인하기 위해, ARIMA

(1,0,0)(1,0,0) 벤치마킹 모형과 독립변수를 추가한 모형을 비교해 보았다. <표 3>을 보면, 웹 검색 트래픽 데이터를 반영한 모형이 벤치마킹 모형보다 R 제곱이 더 높고, RMSE가 더 낮아 설명력이 더 좋다는 것을 확인할 수 있다. 세계의 웹검색 트래픽 변수 중 오직 ‘평창 불거리’만이 유의한 변수로 선정되었고, 평창 여행은 ‘먹거리’나 ‘숙소’, ‘축제’보다 ‘불거리’에 대한 검색 트래픽 변화가 여행수요 변화를 가장 잘 반영하고 있다고 유추할 수 있다.

IV. Conclusion

본 연구는 여행수요를 설명하고 예측하는 데 있어, 웹검색 트래픽 정보가 소비자의 관심을 반영하는 의미 있는 변수로서 중요한 역할을 하고 있음을 보여주고 있다. 대한민국 대표 관광 고시 평창을 중심으로 여행수요 예측모형을 개발하고, 기존의 시계열 모형을 보다 더 우수함을 검증하였다. 구체적으로 웹검색 트래픽 데이터는 소비자 관심을 대변할 수 있고, 여행수요에 시차를 두고 높은 상관관계를 가지며, 여러 키워드 중 여행수요를 잘 설명하는 키워드를 찾는 분석을 통해 지역을 대표하는 특징이 무엇인지 확인할 수 있게 한다.

본 연구는 다양한 측면에서 의미 있는 시사점을 가진다. 첫째, 소비자의 관심을 보여주는 웹검색 트래픽 데이터를 활용하여 여행수요를 예측한 선구적인 연구라는 점이다. 여행수요 예측 연구에서는 주로 시계열 모델링 기반의 진보된 모형을 개발하는 방향으로 연구가 되고 있는데, 본 연구는 여행수요 예측에 의미 있는

새로운 변수로서, 검색키워드 기반 트래픽 데이터가 활용될 수 있음을 보여주고 있다. 두 번째, 웹검색 트래픽 연구 분야에서도 실제 입장객수 기반으로 도출한 여행수요 설명 및 예측에 트래픽 데이터가 활용될 수 있음을 보여주어, 검색 트래픽 활용분야를 넓힌 데 의미가 있다. 세 번째, 정성조사를 정량 모형 개발 전에 체계적으로 도입하여 연구자 주관에 개입될 수 있는 여지를 최소화 하였고, 이는 태생적으로 웹검색 트래픽 연구가 가지는 한계를 일부 극복할 수 있는 방법론이 될 수 있다고 생각된다. 본 연구에서는 연구 대상 도시 선정, 키워드 선정, 시차 확인 등의 정량모형 개발에 필요한 많은 결정 사항들을 정성조사를 통해 소비자 인사이트 기반으로 도출한 점에서 방법론적으로 발전한 연구라 할 수 있다. 네 번째, 실무에서 큰 비용을 쓰지 않고 간단한 분석을 통해, 여행수요를 과학적으로 도출할 수 있는 실질적인 Best Practice를 제안한 점이다. 관련 업체나 기관들이 인터넷 검색서비스 업체와 전략적인 제휴를 통해, 좀 더 정교한 데이터와 키워드를 제공받을 경우, 본 연구에서 보여준 설명력이나 예측력보다 더 우수한 모형을 개발할 수 있을 것이다.

한 국가의 관광정책이란 여행의 주체인 일반 소비자, 여행 대상인 자원과 시설, 그리고 여행 관련 조직체 등을 목표, 환경, 전략 및 효과 등의 요인에 기반하여 최적으로 자원 배분하고 운용하는 일련의 활동이라고 말할 수 있다. 이런 정책 수립의 가장 기본이 되는 데이터는 정확한 여행수요 예측일 것이다. 본 연구가 제안하는 모형은 간단하고 실무 정책자들이 활용하기에 쉬워서, 예측 관련 고급지식이 부족하더라

도 활용할 수 있다고 판단된다. 2018년 1월부터 개최되는 평창 동계올림픽에 대한 수요는 지금까지 경험하지 않은 축제이기 때문에 자기 시계열 데이터에서 수요증가분을 찾아내기는 쉽지 않을 것이다. 이런 경우, 웹검색 트래픽이 반영된 예측모형은 온라인상의 소비자 관심 정도를 반영할 수 있기 때문에, 기존 모형들 대비 더 적합할 것으로 생각된다.

본 연구는 다음의 세 가지 한계점을 갖는다. 첫 번째, 웹검색 트래픽 데이터와 여행수요 데이터 간의 주기가 맞지 않아 연구자가 가공을 했다는 점이다. 두 번째, 소수로 한정하여 검색 키워드의 수를 제한한 점이다. 정성조사 결과 기반으로 키워드를 추출하였으나, 소비자들의 다양한 검색 행태를 모두 포괄할 수는 없었다. 마지막으로 본 연구는 국내수요만 대상으로 한 점이다. 최근 외국인 관광객이 많이 국내로 유입되고 있는데, 데이터 수집의 한계로 인해 국내 여행자들만 고려한 점이 아쉬운 점이다. 외국인 관광수요 데이터들이 체계적으로 수집되고, 본 연구의 방법처럼 해외 유명 검색서비스의 웹검색 트래픽 데이터를 추출할 수 있다면, 조금 더 완전한 예측모형 개발이 가능할 것이다.

참고문헌

강미희, 김남조, 최승담, “대중관광객과의 비교를 통한 생태관광객의 차별적 특성 규명: 여행 동기 및 태도를 중심으로,” 대한관광경영학회, 제17권, 제2호, 2002, pp. 157-175.

강태영, 박도형, “전문가 제품 후기가 소비자 제품 평가에 미치는 영향,” 지능정보연구, 제22권, 제1호, 2016, pp. 63-82.

김범수, 최유지, 박도형, “웹검색트래픽 정보를 활용한 투자모형 개발: KOSPI 지수를 중심으로,” *Entrue Journal of Information Technology*, 제14권, 제3호, 2015, pp. 63-81.

나운중, 박은희, “IT의 기존 관광산업에 대한 영향 및 그 대안으로서의 관광목적지 마케팅시스템에 대한 연구,” 한국관광학회 제50차 학술심포지엄 학술연구발표논문집, 2001, pp. 287-296.

송근석, 송학준, “국제관광수요 결정요인에 관한 연구-방한 일본인을 중심으로,” 호텔관광연구, 제21권, 단일호, 2006, pp. 112-127.

이동원, 강태영, 박도형, “소비자 구매의사결정 단계에 따른 PC 기반 쇼핑과 모바일 기기 기반 쇼핑 행태 차이에 관한 연구: 과업 기술 적합 이론을 중심으로,” *Entrue Journal of Information Technology*, 제13권, 제3호, 2014, pp. 107-122.

진승표, 박도형, “웹검색 트래픽 정보를 활용한 지능형 브랜드 포지셔닝 시스템,” 지능정보연구, 제19권, 제3호, 2013, pp. 93-111.

Butler, D, “When Google Got Flu Wrong,” *Nature*, Vol. 494, No. 7436, 2013, pp. 155.

Choi, H., and Varian, H, “Predicting the Present with Google Trends,” *Economic Record*, Vol. 88, No. s1, 2012, pp. 2-9.

- Crompton, J.L, "Motivations for Pleasure Vacation," *Annals of Tourism Research*, Vol. 6, No. 4, 1979, pp. 408-424.
- Dharmaratne, G.S, "Forecasting Tourist Arrivals in Barbados," *Annals of Tourism Research*, Vol. 22, No. 4, 1995, pp. 804-818.
- Frechtling, D, *Forecasting Tourism Demand: Methods and Strategies*, Butterworth-Heinemann: Oxford, 2001.
- Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., and Brilliant, L, "Detecting Influenza Epidemics Using Search Engine Query Data," *Nature*, Vol. 457, No. 7232, 2009, pp. 1012-1014.
- Jun, S.-P., and Park, D.-H, "Visualization of Brand Positioning Based on Consumer Web Search Information: Using Social Network Analysis," *Internet Research*, Vol. 27, No. 2, 2017, pp. 381-407.
- Jun, S.-P., and Park, D.-H, "Consumer Information Search Behavior and Purchasing Decisions: Empirical Evidence from Korea," *Technological Forecasting and Social Change*, Vol. 107, 2016, pp. 97-111.
- Jun, S.-P., Park, D.-H., and Yeom, J, "The Possibility of Using Search Traffic Information to Explore Consumer Product Attitudes and Forecast Consumer Preference," *Technological Forecasting and Social Change*, Vol. 86, 2014, pp. 237-253.
- Lui, C., Metaxas, P.T., and Mustafaraj, E, "On the Predictability of the US Elections through Search Volume Activity," *Proceedings in e-Society Conference*, 2011.
- Oh, C.O, "The Contribution of Tourism Development to Economic Growth in the Korean Economy," *Tourism management*, Vol. 26, No. 1, 2005, pp. 39-44.
- Park, D.-H., and Kim, S, "The Effects of Consumer Knowledge on Message Processing of Electronic Word-of-mouth via Online Consumer Reviews," *Electronic Commerce Research and Applications*, Vol. 7, No. 4, 2008, pp. 399-410.
- Park, D.-H., and Lee, J, "eWOM Overload and its Effect on Consumer Behavioral Intention Depending on Consumer Involvement," *Electronic Commerce Research and Applications*, Vol. 7, No. 4, 2008, pp. 386-398.
- Park, D.-H., Lee, J., and Han, I, "The Effect of on-line consumer reviews on Consumer Purchasing Intention: The Moderating Role of Involvement," *International Journal of Electronic Commerce*, Vol. 11, No. 4, 2007, pp. 125-148.
- Sheldon, P.J., and Var, T, "Tourism Forecasting: a Review of Empirical Research," *Journal of Forecasting*, Vol.

- 4, No. 2, 1985, pp. 183-195.
- Sims, C.A, "Macroeconomics and Reality," *Econometrica: Journal of the Econometric Society*, Vol. 48, No. 1, 1980, pp. 1-48.
- Song, H., Witt, S.F., and Jensen, T.C, "Tourism Forecasting: Accuracy of Alternative Econometric Models," *International Journal of Forecasting*, Vol. 19, No. 1, 2003, pp. 123-141.
- Song, H., and Witt, S.F, "Forecasting International Tourist Flows to Macau," *Tourism Management*, Vol. 27, No. 2, 2006, pp. 214-224.
- Um, S, "Development of Pleasure Travel Attitude Dimension," *Annals of Tourism Research*, Vol. 18, No. 3, 1991, pp. 500-504.
- Uysal, M., and Crompton, J.L, "An Overview of Approaches Used to Forecast Tourism Demand," *Journal of Travel Research*, Vol. 23, No. 4, 1985, pp. 7-15.
- Vosen, S., and Schmidt, T, "Forecasting Private Consumption: Survey Based Indicators vs. Google Trends," *Journal of Forecasting*, Vol. 30, No. 6, 2011, pp. 565-578.
- Witt, S.F., and Witt, C.A, "Forecasting Tourism Demand: A Review of Empirical Research," *International Journal of Forecasting*, Vol. 11, No. 3, 1995, pp. 447-475.

박도형 (Park, Do-Hyung)



KAIST 경영대학원에서 MIS 전공으로 석사/박사학위를 취득하였다. 현재 국민대학교 경영대학 / 비즈니스 IT 전문대학원 부교수로 재직 중이며, 한국과학기술정보연구원(KISTI)에서 유망아이템 발굴, 기술가치 평가 및 로드맵 수립, 빅데이터 분석 등을 수행하였고, LG전자에서 통계, 시선/뇌파 분석, 데이터 마이닝 등 활용한 연구 및 소비자 평가 모형 개발을 담당하였고, 스마트폰, 스마트TV, 스마트 Car 등에 대한 Technology, Business, Market Insight 기반 컨셉 도출 프로젝트를 다수 수행하였다. 현재 주요 관심분야는 SNS나 온라인 구전 등의 사용자 행동 이론(User Behavior), 사용자 경험 디자인 프로세스 및 혁신 제품 발굴(User eXperience), 빅데이터 기반 사용자 분석(User Analytics) 등이다.

<Abstract>

The Development of Travel Demand Nowcasting Model Based on Travelers' Attention: Focusing on Web Search Traffic Information

Park, Do-Hyung

Purpose

Recently, there has been an increase in attempts to analyze social phenomena, consumption trends, and consumption behavior through a vast amount of customer data such as web search traffic information and social buzz information in various fields such as flu prediction and real estate price prediction. Internet portal service providers such as google and naver are disclosing web search traffic information of online users as services such as google trends and naver trends. Academic and industry are paying attention to research on information search behavior and utilization of online users based on the web search traffic information. Although there are many studies predicting social phenomena, consumption trends, political polls, etc. based on web search traffic information, it is hard to find the research to explain and predict tourism demand and establish tourism policy using it. In this study, we try to use web search traffic information to explain the tourism demand for major cities in Gangwon-do, the representative tourist area in Korea, and to develop a nowcasting model for the demand.

Design/methodology/approach

In the first step, the literature review on travel demand and web search traffic was conducted in parallel in two directions. In the second stage, we conducted a qualitative research to confirm the information retrieval behavior of the traveler. In the next step, we extracted the representative tourist cities of Gangwon-do and confirmed which keywords were used for the search. In the fourth step, we collected tourist demand data to be used as a dependent variable and collected web search traffic information of each keyword to be used as an independent variable. In the fifth step, we set up a time series benchmark model, and added the web search traffic information to this model to confirm whether the prediction model improved. In the last stage, we analyze the prediction

models that are finally selected as optimal and confirm whether the influence of the keywords on the prediction of travel demand.

Findings

This study has developed a tourism demand forecasting model of Gangwon-do, a representative tourist destination in Korea, by expanding and applying web search traffic information to tourism demand forecasting. We compared the existing time series model with the benchmarking model and confirmed the superiority of the proposed model. In addition, this study also confirms that web search traffic information has a positive correlation with travel demand and precedes it by one or two months, thereby asserting its suitability as a prediction model. Furthermore, by deriving search keywords that have a significant effect on tourism demand forecast for each city, representative characteristics of each region can be selected.

Keyword: Travel Demand, Tourist Demand, Web Search Traffic, Naver Trend, Big Data

* 이 논문은 2017년 8월 24일 접수, 2017년 9월 15일 1차 심사, 2017년 9월 25일 게재 확정되었습니다.