

## 웹검색 트래픽 정보를 활용한 유커 인바운드 여행 수요 예측 모형 및 유커마이닝 시스템 개발\*

최유지

국민대학교 비즈니스IT전문대학원  
(uuzie13@gmail.com)

박도형

국민대학교 경영대학/비즈니스IT전문대학원  
(dohyungpark@kookmin.ac.kr)

최근 독감 예측이나 당선인 예측, 구매 패턴, 투자 등 다방면에서 웹검색 트래픽 정보, 소셜 네트워크 내용 등 거대한 데이터를 통해 사회적 현상, 소비 패턴을 분석하는 시도가 이전보다 늘어났다. 구글, 네이버, 바이두 등 인터넷 포털 업체들의 웹검색 트래픽 정보 공개 서비스와 함께 웹검색 트래픽 정보를 활용하여 소비자나 사용자와 관련된 연구가 실시되기 시작했다. 웹검색 트래픽 정보를 활용한 사회 현상, 소비 패턴 분석을 연구는 많이 수행되었으나, 그에 비해서 도출된 여행 수요 모형을 토대로 의사결정을 위한 실질적 대책 수립으로 이어지는 연구는 많이 진행되지 않은 실정이다. 관광산업은 상대적으로 많은 고용을 가능하게 하고 외자를 유치하는 등 고부가가치를 창출하여 경제 전체에 선순환 효과를 일으키는 중요한 산업이다. 그 중에서도 국내 입국 외래객중 수년간 2위와의 큰 차이로 1위를 차지해왔던 중국 국적의 관광객 ‘유커’ 및 그들이 지출하는 1인당 평균 관광 수지는 한국 경제에 매우 중요한 한 부분이다. 관광 수요의 예측은 효율적인 자원 배분과 합리적인 의사 결정에 있어서 공공부문 및 민간부문 모두 중요하다. 적절한 관광 수요 예측을 통해서 한정된 자원을 더욱 효과적으로 활용하여 더욱 많은 부가가치를 창출하기 위한 것이다. 본 연구는 중국인 인바운드를 예측하는 방법에 있어, 이전보다 더 최신의 트렌드를 즉각적으로 반영하고 개인들의 집합의 관심도가 포함되어 예측 성능이 개선된 방법을 제안한다. 해외여행은 고관여 소비이기 때문에 잠재적 여행객들이 입국하기 전 웹검색을 통해 적극적으로 자신의 여정과 관련된 정보를 취득하기 위한 활동을 한다. 따라서 웹검색 트래픽 수치가 중국인 여행객의 관심정도를 대표할 수 있다고 보았다. 중국인 여행객들이 한국 여행을 준비하는 단계에서 검색할만한 키워드를 선정해 실제 중국인 입국자 수와 상관관계가 있음을 검증하고자 하였다. 중국 웹검색 엔진 시장에서 80%의 점유율을 가지는 중국 최대 웹검색 엔진 ‘바이두’에서 공개한 웹검색 데이터를 활용하여 그 관심 정도를 대표할 수 있을 것이라 추정했다. 수집에 필요한 키워드의 선정 단계에서는 잠재적 여행객이 여정을 계획하고 구체화하는 단계에서 일반적으로 검색하게 되는 키워드 후보군을 선정하였다. 키워드의 선정에는 중국 국적의 잠재적 여행객 표본과의 인터뷰를 거쳤다. 트래픽 대소 관계 확인 결과에 따라서 최종 선정된 키워드들을 한국 여행이라는 주제와 직접적인 연관을 가지는 키워드부터, 간접적인 연관을 가지는 키워드까지 총 세 가지 레벨의 카테고리로 분류하였다. 분류된 카테고리 내의 키워드들은 바이두가 제공하는 웹검색 트래픽 데이터 제공 서비스 ‘바이두 인덱스’를 통해 웹검색 트래픽 데이터를 수집했다. 공개된 데이터 페이지 특성을 고려한 웹 크롤러를 직접 설계하여 웹검색 트래픽 데이터를 수집하였고, 분리되어 수집된 변수에는 필요한 변수 변환 과정을 수행했다. 자동화 수집된 웹검색 트래픽 정보들을 투입하여 중국 여행 인바운드에 대한 유의한 영향 관계를 확인하여 중국인 여행객의 한국 인바운드 여행 수요를 예측하는 모형을 개발하고자 하였다. 정책 의사결정 및 관광 경영 의사결정 같은 실무적 활용을 고려하여 각 변수의 영향력을 정량적으로 설명할 수 있고 설득이 명료한 방법인 다중회귀분석방법을 적용해 선형 식을 도출하였다. 수집된 웹검색 트래픽 데이터를 기존 검증된 모형 독립변인들에 추가적으로 투입함으로써 전통적인 독립변인만으로 구성된 연구 모형과 비교하여 가장 뛰어난 성능을 보이는 모형을 확인하였다. 본 연구에서 검증하려는, 웹검색 트래픽으로 대표되는 독립변인을 투입한 최종 도출된 모형을 통해 중국인 관광 수요를 예측할 때 유의한 영향을 끼치는 웹검색 트래픽 변수를 확인

\* 이 논문은 2017년 대한민국 정부(교육부)와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2017R1C1B5016712).

할 수 있다. 최적 모형 설명력을 가지는 모형을 기반으로 최종 회귀 식을 만들었고 이를 ‘유커마이닝’ 시스템 내부에 도입하였다. 데이터 분석에서 더 나아가 도출된 모형을 직관적으로 시각화하고, 웹검색 트래픽 정보를 활용하여 도출할 수 있는 인사이트를 함께 보여주는 데이터 분석 기반의 ‘유커마이닝’ 솔루션의 시스템 알고리즘과 UX를 제안하였다. 본 연구가 제안하는 모형과 시스템은 관광수요 예측모형 분야에서 웹검색 트래픽 데이터라는 정보 탐색을 하는 과정에 놓인 개인들의 인터랙티브하고 즉각적인 변수를 활용한 새로운 시도이다. 실무적으로 관련 정책결정자나 관광사, 항공사 등이 활용 가능한 실제적인 가치를 가지고, 정책적으로도 효과적인 관광 정책 수립에 활용될 수 있다.

**주제어** : 중국인, 유커, 관광수요, 여행 수요, 웹검색 트래픽 정보, 바이두, 빅데이터, 인바운드

.....  
 논문접수일 : 2016년 9월 1일    논문수정일 : 2017년 6월 28일    게재확정일 : 2017년 6월 28일  
 원고유형 : 일반논문            교신저자 : 박도형

## 1. 서론

대한민국은 경제 성장과 동시에 한류를 포함한 다양한 콘텐츠를 앞세워 관광산업의 규모가 커지고 있다. 실제로 국가 통계에 따르면, 외국인 입국자 수가 2011년 9,794,956명에서 2014년에는 14,201,516명으로 약 1.5배 증가하였고, 2013년 국내 GDP에 대한 관광산업이 미치는 효과가 698억 달러로 성장한 수치를 보였다. 이렇게 커지는 관광산업을 흔히 황금알을 낳는 거위라고 표현하는데 그 이유는 외화 수입을 통해 국제수지가 개선되고 경제가 활성화 되고, 내수 경기가 살아나며 부가적인 일자리 창출과 고용 효과가 따라오기 때문이다. 관광 지식정보시스템의 조사에 따르면 2015년 국내 고용시장에서의 관광사업 비율은 전체 고용시장에서 6.2%로 예측되고 있을 만큼 관광산업의 부가적인 가치를 무시할 수 없다.

특히 그 중에서도 2010년 G2로 부상한 중국을 제외할 수 없다. 중국은 해외여행객이 2001년 1,213만 명에서 2014년 1억 900만 명으로 증가할 만큼 세계관광시장의 핵심이 되고 있으며 많은 국가들이 중국 해외여행을 잡기위해 앞 다투어

유치에 힘쓰고 있다. 일본은 규제를 완화하고 관광패스라인을 구축하는 등의 노력을 하고 있지만 지리적으로 가까우며 한류 열풍을 발판삼아 유커(Yóukè)를 잡기위해 다양한 정책을 펼치는 한국에 밀려, 중국 아웃바운드 대상국에서 홍콩, 마카오, 한국으로 3위의 자리를 내어주고 말았다 (Jang, 2015). 문화체육관광부 제공 출입국관광통계 데이터를 확인해보면, 중국 국적 외래객의 한국 입국 숫자는 2010년 187만 명에서 2014년 612만 명, 2016년 806만 명으로 계속 늘어나는 추세를 보였다. 한국관광공사 외래 관광객 실태조사 데이터를 확인해 보면 중국인 관광객 1인당 평균 지출금액이 15년 기준 2,483 달러까지 다년간 꾸준히 증가하는 추세를 보인 만큼, 그 영향력을 무시할 수 없다.

관광산업이 한국 경제에 미치는 막대한 선순환적 효과와 중국인 관광객 유치가 중요한 상황속에서, 얼마나 많은 관광객이 한국에 들어올 것인가에 대한 수요예측의 필요성과 중요성을 생각해볼 수 있다. 예를 들어, 정확한 수요예측은 보다 효율적인 항공편 좌석 관리로 효율적인 수용력과 지역 산업에 도움을 줄 수 있다. 이렇듯, 적절한 수요예측은 자원의 효율적인 관리로 수

의 극대화의 기회 및 성공가능성 향상을 위한 의사결정을 도와줄 수 있을 것이다.

본 연구는 보다 정확한 예측과 신속한 정보 반영을 위해, 과거의 데이터 및 소비자 행동에 기반을 두어 수집한 데이터들을 이용해 수요 예측 분석을 실시하였다. 예측력 개선의 일환으로 소비자 개인의 관심을 측정하기 위해 웹검색 트래픽 정보를 수집하였다. 빅데이터 시대의 개막과 함께 인터넷 사용자들의 웹검색 트래픽 정보가 공개되고 이를 활용하여 전통적인 방식보다 개선된 예측모형을 세운 연구가 다방면에서 활발히 진행 중이다(Jun and Park, 2017). 웹검색 트래픽 정보는 소비자 동향, 물가, 부동산, 독감, 금융투자 등 다양한 분야에서 활용되고 있다(Jun and Park, 2013; Jun et al. 2014; Kim et al. 2015; Jun and Park, 2016). 지리적으로 멀리 떨어진 곳을 방문하여 여러 경험을 갖게 되는 여행이나 관광의 경우, 실제 소비자들은 준비단계에서 인터넷 웹 정보 탐색을 통해 다양한 정보를 찾고 구체적인 계획을 세우기 때문에(Park and Lee, 2015), 웹검색 트래픽 정보가 여행 수요 예측에 매우 유용할 것으로 예상할 수 있다. 하지만 현재까지 웹검색 트래픽 정보가 여행 수요 예측에 활용될 수 있음을 보여준 연구는 전무한 실정이다.

본 연구는 소비자의 행동 기반 산출물인 웹검색 트래픽 정보를 바탕으로 여행 수요 예측 모형을 개발하고 모형의 우수성을 검증하고자 한다. 기존의 예측 모형이 다른 변수에서 추가적으로 웹검색 트래픽 정보가 모형에 고려될 경우 우수함을 보여줄 것이다. 또한, 본 연구에서는 연구에서 세워진 모형에 따른 수요예측 결과를 직관적인 그래픽 형태로 보여주는 기능을 포함한 시스템 ‘유커마이닝’을 구현하였다. 수요예측 시스템 ‘유커마이닝’을 통하여 주먹구구식 대안이 아

닌 과학적인 의사결정을 뒷받침해주는 웹클라이언트 솔루션이다. 연구 설계 시에도 실제 정책결정 상황을 고려한 분석방법을 채택하는 등 전반적인 과정이 실질적인 활용을 염두에 둔 채로 진행되었다. 최종적으로는 전통적인 방식과 비교했을 때 상대적으로 적중률이 높은 예측 식을 세워 실질적인 의사결정 과정에 활용되는 모형과 시스템이 되기를 기대한다.

## 2. 이론적 배경

### 2.1 관광산업에서의 수요 예측

관광산업의 성장은 고용율과 더 나아가 GDP에 직접적, 간접적으로 미치는 효과가 뚜렷하다. 또한 내수 증진, 외자 유치 등 경제적 파급 효과도 무척 크다(Lee et al., 2006). 관광산업의 지속적인 성장과 경쟁우위를 위하여 신속, 합리적인 의사결정이 중요하다. 앞서 말한 항공좌석 예측에 수요 예측이 필요한 것 외에도, 호텔 객실 등의 분야에서 만일 과소 예측하는 경우 물리적인 수용 능력 부족, 노동력 부족 등 때문에 수입 기회를 상실할 수도 있다. 관광시설 또한 수용력 초과 시 혼잡성이 증가하여 고객의 불평을 초래할 수도 있다. 관광 수요예측 연구는 수요 예측을 정확하게 하는 것이 공공부문과 민간부문의 더 나은 의사결정을 위해 중요하다는 것을 강조한다(Uysal and Crompton, 1985; Sheldon and Var, 1985; Witt and Witt, 1992, 1995; Dharmaratne, 1995; Archer et al., 1996; Song et al., 2003). 관광수요예측 방법은 크게 세 가지로 나뉜다. 수치를 그대로 이용하는 정량적 방법과 주관적 판단을 개입시키는 정성적 방법, 이

두 가지를 적절하게 결합한 방법이다.

Song et al.(2003)의 연구에서는 관광 수요 예측이 지방정부의 정책입안자에게도 중요한 부분이고 높은 정확성과 간단한 모형이 중요하다고 했다. Archer et al.(1987)은 마케팅, 생산, 회계 등 전사적인 방면의 관광업 기업체의 전략에서 손실을 최소화하기 위한 정확한 관광수요 예측의 필요성을 강조한 연구를 출판했다. Ahn et al.(1995)은 한국의 외래 관광객, 국내·국외관광에 대한 관광수요예측 모형을 세우는 연구에서 정량적 예측방법과 정성적 예측 방법으로 분석, 두 가지를 복합적으로 적용 하였다. 정량적 예측방법에는 시계열 분석법, 인과관계 분석법, 정성적 예측방법에는 델파이 예측 기법, 시나리오 설정법이 사용되었다. Lee and Jung(2004)은 시계열 분석 기법들을 이용하여 월드컵 기간의 외래 관광객 입국자수 예측연구의 적절도에 대하여 연구하고 2002년 월드컵기간 입국자수를 예측하였다. Jeong et al. (2013)은 중국 인바운드 관광수입을 예측하였는데, ARIMA 모형을 통하여 2012년 \$502.15억, 2013년 \$534.45억, 2014년 \$566.01억, 2015년 \$597.57억, 2016년 \$629.12억의 관광수입이 창출될 것으로 예측했다. 기존 분석의 경우, 연 기준으로 쌓이는 데이터를 분석하여 해당 연도 및 다음 연도를 예측하는 모델을 세우는 연구들이었다.

## 2.2 웹검색 트래픽 정보 관련 선행연구

정보기술 분야뿐 아니라 다양한 분야에서 빅데이터에 대한 관심이 상승함에 따라(Kang and Park, 2016), 검색엔진 사이트들도 구글을 선두로, 검색어 당 유저들의 웹검색 트래픽 정보를 공개하기 시작했다. 웹검색 트래픽 정보 데이터

를 활용한 연구의 장점은 최근 유저들이 관심을 갖고 검색해본 검색추이를 통해 가까운 미래를 예측하는 데에 탁월하다(Ginsberg et al., 2009). 현재 구글 트렌드, 네이버 트렌드, 바이두 인덱스 등의 서비스를 제공하고 있다. 각 검색엔진에서 검색한 횟수를 상대 및 절대 수치로 알아볼 수 있으며 검색엔진을 이용하는 유저들의 관심도를 시간별로 확인해볼 수 있다. 접근성이 높고 획득하기 쉬운 웹검색 트래픽 정보데이터 제공 서비스의 등장에 따라 해당 데이터를 활용한 연구도 활발해졌다. 웹검색 트래픽 정보는 제조, 마케팅, 서비스, 금융, 보건정책 등 다양한 방면에서 연구에 활용되었으며 탄력재와 비탄력재, 유동성을 막론한 분야에서 주목을 받고 있다.

웹검색 트래픽 정보가 예측에 효과적으로 사용되는 것을 증명한 것이 본격적으로 알려지기 시작한 연구가 Ginsberg et al.(2009)의 독감 바이러스 징후 예측 논문이다. 구글은 2008년 “독감 트렌드(Flu Trends)” 서비스를 공개하였는데, 사람들이 독감에 걸렸다고 생각할 때 검색할만한 키워드로 독감바이러스 확산에 대하여 예측하는 서비스이다. 이 서비스를 통하여 해당 연구에서 미국 질병통제센터 CDC보다도 1~2주 빠르게 독감 바이러스 유행을 감지할 수 있었다. Butler(2013)가 이 연구에 대해 의구심을 제기 했으나 웹검색 트래픽 정보 데이터가 가지는 신속성과 편리성이 활용도 높은 정보로서의 가능성을 보였다.

Choi and Varian(2012)은 구글 트렌드가 제공하는 웹검색 트래픽 정보를 활용해 단기적 예측에 더 나은 성과를 증명했다. 경제활동을 규모별로 크게 자동차 판매, 주택판매, 소매 세 분류로 나누었다. 경제적 시간에 따른 수요의 변곡점을 알게 해주는 측면에서 장기적 예측보다는 가까

은 단기적 미래를 예측하는 것에 효과적임을 주장한다. Lui et al(2011)은 2008년, 2010년 미국 의회 선거 당시 구글 트렌드 웹검색 트래픽 정보를 활용해 당선자를 예측하여 웹검색 트래픽 정보가 다양한 도메인 분석에 적용됨을 시사했다. Vosen and Schmidt(2011)는 구글 트렌드 웹검색 트래픽 정보와 미시건 소비자 심리지수 MCSI (University of Michigan Consumer Sentiment Index)와 소비자 신뢰지수(Consumer Confidence Index) 각 요인을 비교하여 웹검색 트래픽 정보의 개선된 예측력을 증명했다.

인터넷 검색트렌드와 주식시장의 관계에 대한 연구를 통해 투자자의 검색엔진을 활용한 투자 대책을 제안한 연구(Koo and Kim, 2015)와 인터넷 검색트렌드를 이용해 주식시장과의 관계를 밝히고 전통적인 방식보다 수익성이 높은 투자 모델을 설계한 연구(Kim et al, 2015)도 있는 등 활발한 연구가 이루어지고 있는 실정이다. 또한 에스앤유 리서치에서 웹검색 트래픽 정보데이터와 SNS등 웹상에서 얻을 수 있는 유저의 관심 빅데이터를 활용한 SPP지표와 함께 초단기 투자자를 위한 종목별 예측, 매수 및 매도 타이밍 가이드 서비스를 제공하고 있다.

소개된 기존 연구들은 다양한 분야에서 사회 현상을 예측하는 데에 웹검색 트래픽 정보가 유용한 정보로 활용될 수 있음을 보여주고 있다. 웹검색 트래픽 정보는 설문이나 면담, 관찰 등의 방법보다 비교적 쉽고 저렴한 방법으로 수집 가능하며 기간별로 즉각적인 반영도 가능하다. 예측하려는 집단을 대표할 수 있는 수집기관을 잘 선정한다면 소비자의 관심도가 반영된 지표로 손색이 없는 데이터이다.

### 2.3 관광 수요예측을 위한 웹검색 트래픽 정보 활용

관광산업에서 수요예측의 중요성은 매우 큰데, 이는 관광업에 직접적으로 종사하고 있는 사람들뿐만 아니라 지역 산업에도 다각적으로 영향을 미치기 때문이다. 관광수요와 관련된 기존 연구들을 살펴보면 주로 연간 데이터를 이용하는 경우가 빈번하다. 연간 데이터를 이용하여 분석 모델을 설계할 경우, 신속하고 정확한 판단이 필요한 의사결정자들에게 실질적인 도움을 주기 어렵다. 기존의 변수만으로는 예기치 않은 상황에 일어나는 갑작스러운 관광수요의 변화를 장기적으로 예측하기 어렵다. 따라서 본 연구는 ‘웹검색 트래픽 정보’라는 실시간성 데이터를 고려하여 관광수요 예측가능성을 검증할 것이다.

웹검색 트래픽 정보를 이용하여 수요예측 모델을 만드는 다양한 분야의 연구가 등장했으나 관광 수요예측모델을 구현하기 위해 웹검색 트래픽 정보를 활용한 경우는 Park and Lee(2015)의 대한민국 강원도 주요 도시에 대한 관광수요 예측 모형을 주목할 수 있다. ARIMA 시계열 분석을 통해 도시별 특징에 따라 다른 키워드가 유의한 영향을 미치는 요인이 되는 것을 확인하면서 개선된 모형을 도출 했다.

본 연구에서는 외국인 국내여행객수요 중에서도 유커라고 불리는 중국인 여행객의 수요를 예측하는 데에 웹검색 트래픽 정보가 유의미한 독립변수로 작용할 것이라는 가설에 대하여 연구를 진행한다. 선행연구들의 관광수요 예측모델에서 유의하다고 검증된 변수들을 벤치마크 모델로 두고 본 연구에서 검증하고자 하는 웹검색 트래픽 정보 변수를 추가하여 모델의 적절성과 예측력을 비교한다. 어떤 키워드에 대한 웹검색

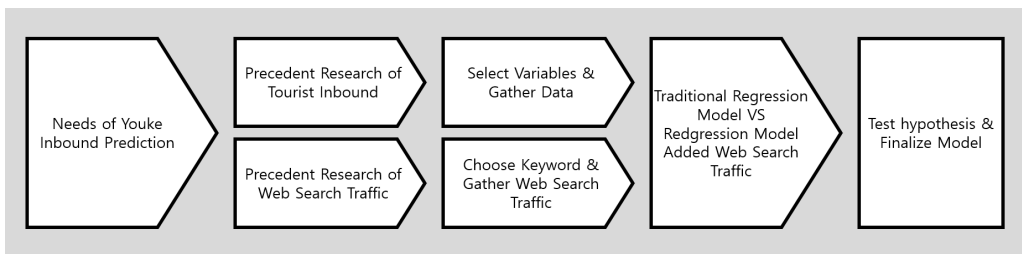
트래픽 정보 데이터를 사용하는 지가 중요하게 작용할 것이다. 따라서 본 연구에서는 한국 여행이라는 주제와 직접적인 연관 정도에 따라 세 카테고리 나눈 단어를 선정했다. 첫 번째는 한국 여행과 직접적으로 관계있는 키워드들로서, 여행 자체를 의미하는 키워드(ex. 한국관광(韩国旅游)), 여행의 행선지(예. 제주도(济州岛)) 등이 사용되었다. 두 번째는 한국여행의 목적과 수단이 되는 한국 여행과 간접적으로 관계있는 키워드들로서, 쇼핑 등의 관광 목적(예. 한국화장품(韩国化妆品))과 사용 언어 등의 여행에 필요한 것(예. 한국어(韩语)) 등이 활용되었다. 마지막으로 세 번째 카테고리는 여행과 직접적으로 관련은 없지만, 한국 여행을 유발하는 한국에 관한 문화적인 트렌드 관련 키워드(예. 한국 드라마/한드(韩剧))를 수집하여 연구 모형에 반영하였다. 잠재관광객이 관광목적지 선택 단계에 따라서 정보를 얻는 양상이 각각 다르다고 할 때 여행을 결정할 때부터, 여행 일정이 구체화 되는 과정. 한국 여행을 하는 중국인 관광객들 중 가장 많은 비중을 차지하는 대표 그룹의 인터뷰를 통하여 중국인 중에서 잠재적 국내여행객을 대표할만한 키워드를 선정했다. 이들 키워드 카테고리들은 유커 여행 수요 예측을 위하여, 여행과 거리가 가까운 키워드부터 거리가 있는 키워드들까지

유커 여행 수요 예측을 잘 설명할 수 있는 키워드들이 무엇인지 확인하는 데 활용되었고, 이는 본 연구의 최적 모형을 도출하는 과정에서 의미 있게 활용될 것이다.

### 3. 연구 설계 및 모형

#### 3.1 연구 설계 및 프로세스

본 연구의 프로세스는 <Figure 1>과 같다. 첫 번째 단계에서는 중국인 관광수요 예측의 필요성에 대하여 기술한다. 두 번째 단계에서는 관광수요 및 웹검색 트래픽 정보 관련 연구들을 두 방향에서 평행하게 문헌 연구를 수행하였다. 세 번째 단계로 중국의 최대 포털 사이트인 바이두에서 어떤 키워드들이 종속변수와 관련 있는지 확인하였다. 네 번째 단계에서는 종속변수로 활용될 중국인 입국자 수 데이터를 수집하고, 독립변수로 활용될 각 키워드의 웹검색 트래픽 정보 및 기존 선행연구에서 이용한 데이터들을 수집하였다. 다섯 번째 단계에서 다중회귀분석을 통하여 모델을 설정하고, 이 모델에 웹검색 트래픽 정보를 추가하여 예측모형이 나아지는 지를 확인하였다. 마지막에는 성능의 우수를 비교하여 최종적으로 도출된 예측모형을 분석하고, 본 연



<Figure 1> Process of Research

구가 검증하고자 하는 웹검색 트래픽 정보와 중국인 관광수요와의 관계, 각 키워드에 따라 영향력이 달라지는 지 확인하였다.

### 3.2 연구 모형 및 변수

본 연구는 전술한 바와 같이 중국인 관광객의 웹검색 트래픽 정보를 활용하여 대한민국으로의 여행객 수요를 예측하기 위한 연구로 선행연구에서 유의하다고 선정된 변수들을 벤치마킹 모델로 하여 진행하였다.

선행연구를 살펴본 결과, 수요예측을 위한 방법은 ARIMA분석, 다중회귀분석, 인공신경망, 패널 분석 등으로 매우 다양하다. 본 연구에서는 의사결정에 실질적으로 활용되는 모형을 목표로 했기 때문에 보다 설득이 명료하고 각 변수의 영향력을 정량적으로 설명할 수 있는 분석 방법인 다중회귀분석을 이용하여 다양한 변수들 간의 회귀 식을 세웠다.

웹검색 트래픽 정보가 수요 예측을 하는 데에 유의미한 변수로 작용할 것이라는 가설 검정을 위하여 다양한 키워드에 대한 트래픽변수를 고려하였다. 웹검색 트래픽 정보 변수가 한국 관광이라는 주제에서 직접적인 정도에 따라 단계별로 분리하여 다양한 회귀모형을 설계하였다. 가설 검증 이외에도 어떤 모델이 예측력이 가장 높은지, 어떤 검색어의 트래픽을 고려했을 때 예측력이 가장 높은지, 어떤 변수들을 고려하였을 때 예측력이 가장 높은지 등을 확인하였다.

중국인 입국자 수 데이터를 종속 변인으로 사용하였다. 해당 데이터는 관측 시점으로부터 약 2개월 후 정리되어 한국관광공사 홈페이지에 게시된다. 독립 변인을 결정할 때 해당 연구 결과인 회귀모형(예측식)의 실증적 활용을 예상하여

최근 데이터를 비교적 신속하게 얻을 수 있는 변수들을 중심으로 수집하였다.

웹검색 트래픽 정보데이터, 국제유가, 소비자물가, 환율, 기온을 독립변인으로 사용하였다. 우선 웹검색 트래픽 정보데이터는 중국 웹검색 엔진 시장의 80%를 차지하는 바이두에서 제공하는 인덱스를 활용하였다. 바이두 인덱스의 장점은 검색 데이터를 일 단위로 주 단위로 확인할 수 있고, 바로 전 날의 트래픽 검색량을 다음 날에 바로 조회할 수 있는 즉시성에 있다. 그러나 구글 트렌드나 네이버 트렌드와 다르게 바이두 인덱스는 일반 사용자에게 웹검색 트래픽 정보데이터를 엑셀 파일로 제공하고 있지 않기 때문에 웹 크롤러를 자체 개발하여 데이터를 획득하였다. 본 연구진이 개발한 웹 크롤러는 JAVA언어를 사용하여 DBMS에 연동하는 방식으로 웹에서 바로 수집하였다. 수집할 키워드를 선정하는 단계에서 일반 소비자층에 해당하는 중국 현지인 표본을 뽑아 인터뷰를 통하여 자문을 구하였다. 바이두와 한국관광공사에 따르면 한국관광과 관련된 키워드를 검색한 키워드를 가장 많이 검색한 인구학적 계층은 30대의 여성으로 확인되었다. 따라서 본 연구에서도 해당 결과를 참고하여 자문 상대를 균형 있게 선정하였다. 또한 홍콩, 대만을 포함 하지 않고 중국 본 대륙의 관광객만을 예측할 것이기 때문에 중국어 간체로 된 키워드를 선택하여 번체 검색어를 고려하지 않았다. 중국 잠재 관광객들과의 인터뷰 결과를 기반으로 실제 트래픽 양을 상호 비교하여 추려진 키워드는 한국관광(韩国旅游), 제주도(济州岛), 한드(韩剧), 제주도여행전략(济州岛旅游攻略) 한국어(韩语), 한국화장품(韩国化妆品) 이었다. 검색어 키워드가 한국관광과 직접적으로 연관된 정도를 파악하여 다음과 같이 세 카테고리

로 나누어 보았다.

카테고리1 - 한국여행과 직접적으로 관련 있는  
키워드: 한국관광(韩国旅游), 제주도(济州岛),  
제주도여행전략(济州岛旅游攻略)

카테고리2 - 한국여행과 간접적으로 관련 있는  
키워드: 한국화장품(韩国化妆品), 한국어(韩语)

카테고리3 - 한국여행을 유발한 트렌드 키워드  
(한류): 한드(韩剧)

한국여행과 직접적으로 연관된 키워드는 물론, 간접적으로 연관된 키워드와 한국관광객 수요의 상관관계도 살펴보는 연구를 통하여 새로운 의미가 있는 결과를 기대하였다. 또한 한류에 대한 관심도를 알아보기 위하여 간접적인 키워드에 분석 기간 동안 관심도가 가장 높았던 상위 9개의 드라마 명에 대한 웹검색 트래픽 정보데이터를 수집한 후, 트래픽 수치의 총합을 계산하여 ‘인기드라마 합’이라는 변수를 생성하였다. 상위 9개의 드라마는 ‘별에서 온 그대(来自星星的你),’ ‘상속자들(继承者们),’ ‘피노키오(匹诺曹),’ ‘옥탑방 왕세자(屋塔房王世子),’ ‘주군의 태양(主君的太阳),’ ‘시티헌터(城市猎人),’ ‘너의 목소리가 들려(听见你的声音),’ ‘청담동엘리스(清潭洞爱丽丝),’ ‘넌 내게 반했어(你为我着迷)’ 이다.

트래픽 외의 독립변인은 선행연구의 증명을 기반으로 하여 정하게 되었다. 우선 국제유가는 정치 및 경제, 국제 외교 상황, 환경 등 사회 전반의 여러 변인들로부터 영향을 받는다. 국제유가의 유동적 변화에 따라서 다양한 산업 및 정책 등 현상이 달라지기도 한다. 본 연구에서

소비자물가지수(CPI: Consumer Price Index)는 소비생활을 영위하는 데 사용하기 위해 구입하

는 상품과 서비스의 가격변동, 추이를 파악하기 위해 작성되었다. 중국인의 아웃바운드에는 대한민국의 소비자물가지수보다 중국의 소비자물가지수가 더욱 큰 영향을 미칠 것으로 예상하여 본 연구는 중국의 소비자물가지수를 사용하였다.

환율은 해외여행에 영향을 미치는 큰 요인 중 하나이다. 환율은 위안-원화 환율을 이용하였다. 환율 데이터는 일, 주, 월, 연간 등 다양한 단위로 여러 기관에서 제공하고 있으며, 본 연구에서는 환율 데이터를 이용하였다.

일 년의 계절성을 반영하는 변수로 기온을 포함하였다. 본 연구는 기상청에서 제공하는 다양한 기온 데이터 중 시기는 매월 중간 값인 15일의 기온이다.

## 4. 연구결과 및 시스템 개발

### 4.1 연구 결과

본 연구가 제안하는 웹검색 트래픽 정보가 유커 여행 수요 예측에 중요한 변수로 활용될 수 있음을 보여주기 위해, 먼저 선행연구에서 제안하고 검증한 변수를 기본으로 이루어진 벤치마크 모형을 개발하였다. 그 다음으로 본 연구가 제안하는 검색 트래픽 정보 변수를 더한 여행 수요 모형의 설명력을 벤치마크 모형과 비교하였다.

중국인 입국자수를 종속변수로, 선행연구에서 사용한 변수들을 독립변수로 설정하고 모형을 개발하였다. <Table 1>을 확인해보면 유가를 제외하고 다른 변수들은 통계적으로 유의한 것으로 판명되었다. 공선성 진단에서도 모든 변수들이 독립적이며 공선성이 문제되지 않는다고 판



〈Table 1〉 Benchmark model(Regression model Based on Conventional Research Variables)

	Unstandardized Beta	Standardized Beta	t-Value	p-Value	VIF
(Constant)	-2577094.041		-3.824	.000	
Chinese CPI	37485.905	.746	8.653	.000	1.586
Currency	-7119.242	-.251	-3.202	.002	1.307
Korean Temperature	5357.033	.335	4.839	.000	1.023
Oil Price	-640.901	-.093	-.975	.334	1.928
<i>Adjusted r<sup>2</sup> = 0.738, F(4,52) = 40.393 (p &lt; 0.001)</i>					

단(VIF<10)하였다. 벤치마크 모형의 수정된 R 제곱은 0.738이다. 후에 트래픽 변수를 추가한 모형의 수정된 R 제곱과 비교할 것이다.

모형 1에서는 선행연구에서 사용한 변수들과 한국여행과 직접적으로 연관된 트래픽 데이터를 이용하여 모형을 설계하였다. 수집된 세 개의 키워드 중 시행착오를 통해 영향력이 큰 변수인 ‘제주도’ 키워드의 웹검색 트래픽 정보를 반영하여 모형을 개발하였다(〈Table 2〉 참조). ‘제주도’ 키워드가 진입하면서 한국기온의 영향력이 낮아

졌지만, 모형 자체의 설명력은 크게 증가하였음을 알 수 있다. 흥미로운 사실은 기존의 연구에서 제안된 변수보다 ‘제주도’ 키워드의 영향력이 지배적이라는 점이다. 중국인 여행자를 설명함에 있어 웹검색 트래픽 정보의 강력함을 보여주는 모형이라 할 수 있겠다.

모형 2에서는 선행연구에서 사용한 변수들과 한국여행과 직접적으로 연관된 트래픽 정보에 추가하여 한국여행과 간접적으로 연관된 트래픽 정보를 포함시켰다. 이번에도 사전에 여러 키워

〈Table 2〉 Model 1 (Including Keyword Directly-Category1- Related with Korean Tourism)

	Unstandardized Beta	Standardized Beta	t-Value	p-Value	VIF
(Constant)	-616602.192		-.949	.347	
Chinese CPI	13088.983	.260	2.310	.025	4.212
Currency	-4276.431	-.150	-2.301	.026	1.419
Korean Temperature	1646.055	.103	1.471	.147	1.624
Oil Price	-421.805	-.061	-.797	.429	1.939
“Jeju Island”	3054.226	.637	5.454	.000	4.519
<i>Adjusted r<sup>2</sup> = 0.831, F(5,51) = 56.124 (p &lt; 0.001)</i>					

(Table 3) Model 2 (Including Keyword Directly and Indirectly-Category2-Related with Korean Tourism)

	Unstandardized Beta	Standardized Beta	t-Value	p-Value	VIF
(Constant)	259927.083		.436	.664	
Chinese CPI	-52.416	-.001	-.009	.993	5.851
Currency	-2907.483	-.102	-1.780	.081	1.474
Korean Temperature	617.137	.039	.621	.537	1.723
Oil Price	977.183	.141	1.746	.087	2.920
“Jeju Island”	2169.885	.452	4.136	.000	5.333
“Korean Cosmetic”	6119.370	.607	4.312	.000	8.836

*Adjusted  $r^2 = 0.874$ ,  $F(6,50) = 66.007$  ( $p < 0.001$ )*

드들 중 시행착오를 통해, 가장 영향력 있는 키워드 ‘한국화장품’을 사전에 선별하였다. <Table 3>을 확인해보면, 한국 여행에 직접 간접적으로 영향 있는 변수들이 포함되어, 기존의 변수들은 0.05 기준에선 유의하지 않다고 판명되고, 웹검색 트래픽 정보만이 유의한 변수로 선정되었다. 모형의 설명력은 소폭 증가한 것을 알 수 있다. 공선성 진단에서도 VIF 수치가 상당부분 상승하였지만, 10이하이기 때문에, 문제가 되는 수준은 아니다. VIF 수치의 변화를 살펴보면, ‘한국화장품’ 웹검색 트래픽 정보는 유가와 상당부분 관계가 있음을 추측할 수 있다. 유가가 감소할수록, 화학제품인 한국화장품의 가격이 하락하여 수요가 증가하게 되는 현상과 밀접한 관련이 있다고 해석할 수 있다. 기존 모형에서 유가를 중요하게 고려하였다면, 이것이 실제 제품에 반영되어 소비자에게 체감되는 유가 관련 검색 키워드가 유가의 변화도 함께 설명할 수 있음을 알 수 있다.

마지막 모형인 모형3에서는 선행연구에서 사

용한 변수들과 한국여행과 직접적/간접적으로 연관된 트래픽 정보에 한류 관련 키워드를 추가하여 모형을 개발하였다. 한국 드라마 9개의 트래픽을 가공하여 만든 ‘한드’ 변수의 영향력을 보여주는 것이 <Table 4>이다. 분석 결과 ‘한드’는 유의한 결과를 보여주지 않았고, 모형3은 모형2보다 낮은 설명력을 보였다. 웹검색 트래픽 정보의 예측력은 먼 시점 예측(Forecasting)보다 근 시점 예측(Nowcasting)에 더 효과적이라는 기존의 연구와 일맥상통한 결과로 추측되는데, 한국드라마 등의 한류의 효과는 즉각적으로 반응하는 것이 아니라 문화 속에 침투하여 중국인 소비자에게 영향을 주는 데 충분한 시간이 필요하기 때문에, 단기 여행 수요를 예측하는 데에는 영향력이 적다고 설명할 수 있을 것이다.

이상의 연구 결과를 통해, 최적의 변수모형을 도출하기 위하여, 0.1 유의수준 안에서 적합한 변수들을 선정하여 최종 모형을 개발하였다. 최적 모형의 설명력은 0.872로서 모형2에 근접하

(Table 4) Model 3 (Including Keyword Directly and Indirectly-Category3-Related with Korean Tourism)

	Unstandardized Beta	Standardized Beta	t-Value	p-Value	VIF
(Constant)	319118.514		.524	.603	
Chinese CPI	-419.908	-.008	-.072	.943	5.925
Currency	-2861.632	-.101	-1.737	.089	1.478
Korean Temperature	701.985	.044	.694	.491	1.762
Oil Price	948.663	.137	1.677	.100	2.943
“Jeju Island”	2202.712	.459	4.145	.000	5.398
“Korean Cosmetic”	6068.479	.602	4.239	.000	8.871
“Korean Drama”	-22.574	-.028	-.565	.574	1.099

*Adjusted r<sup>2</sup> = 0.873, F(7,49) = 55.853 (p < 0.001)*

며, 사용된 변수는 ‘제주도’ 웹검색 트래픽 정보 ( $t = 5.693, p < 0.01$ )과 ‘한국화장품’ 웹검색 트래픽 정보( $t = 4.967, p < 0.01$ ), 마지막으로 환율( $t = -2.826, p < 0.01$ )이다. 세 개의 변수만으로 중국인 입국자 통계 분산의 87.2%를 설명할 수 있었고, 각각의 비표준화 계수를 기반으로 회귀 식을 만들어보면 다음과 같다.

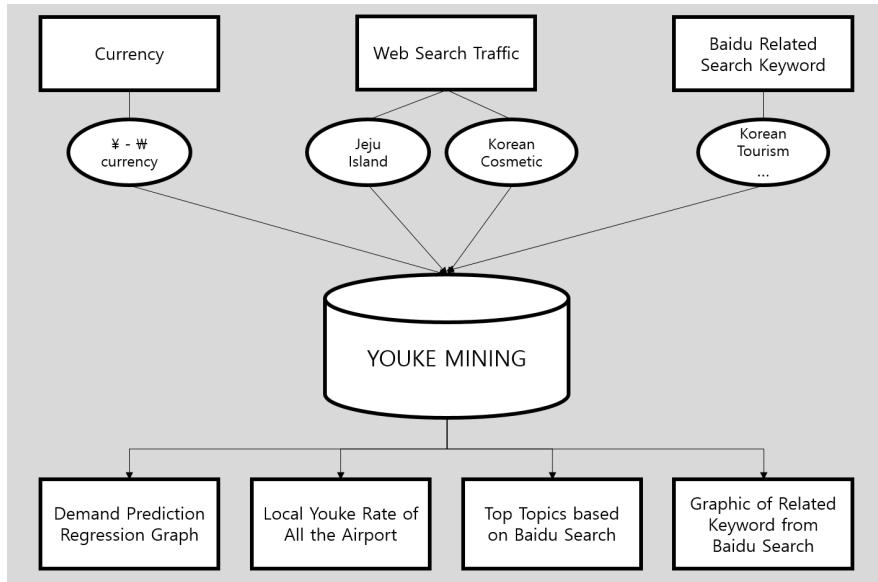
$$\text{중국인 입국자 수} = 588219 + 2508 * \text{‘제주도’ 검색건수} + 4707 * \text{‘한국화장품’ 검색건수} - 4020 * \text{위안/원화 환율}$$

이상의 최종 회귀 식은 유커마이닝 시스템의 수요예측 기능의 모형 식으로 사용되었다.

#### 4.2 시스템 개발

본 연구의 모형을 토대로 하여 관광업 정책결정자의 의사결정을 뒷받침해주기 위한 시스템을

개발하였다. 통계 툴을 다룰 수 없는 사람도 본 연구의 예측 식을 이용하여 관광수요를 예측할 수 있도록 하는 솔루션을 제공한다. 해당 솔루션은 UX트렌드를 반영하여 단일 페이지로 구성하여 깔끔하고 부드러운 사용자 경험을 제공한다. 사용자는 별도의 설치 없이 웹에 접속하여 바로 서비스를 이용할 수 있는 Thin-client 방식이다. ‘유커마이닝’은 잠재적 유커와 관련하여 의사결정시 유용하게 볼 수 있는 서비스를 다양하게 제공하며 큰 특징은 크롤링 기법을 이용하여 웹으로부터 즉시 필요한 정보를 수집하여 기능에 따라 다른 방법으로 화면에 출력하는 것이다. ‘유커마이닝’은 ‘VARIABLE FORECASTING’:수요 예측 시계열 그래프, AIRPORT FORECASTING’: 전국 공항 유커 비율 시각화, ‘TRAFFIC FORECASTING’:바이두 트래픽 기반 상위 콘텐츠의 시각화 재구성, ‘MIND MAP’:바이두 연관 검색어 마인드맵 재구성으로 네 가지의 주요 기



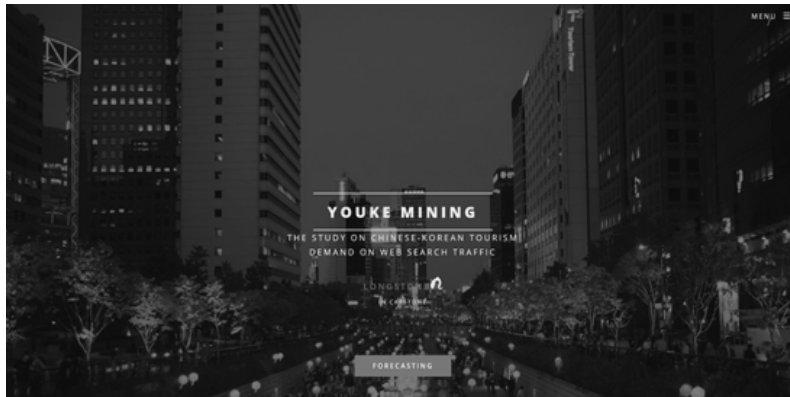
<Figure 2> System Design for 'Youke Mining'

능을 제공한다. ‘유커마이닝’의 시스템 구조는 <Figure 2>와 같다.

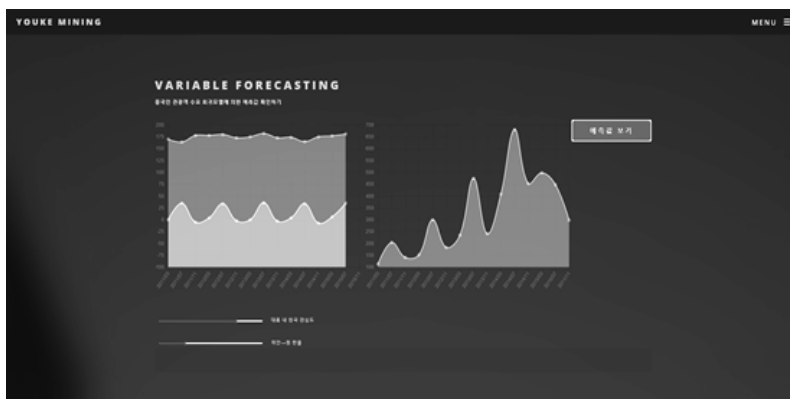
‘VARIABLE FORECASTING’의 경우에는 위안-원화 환율, ‘제주도’ 검색 트래픽 정보, ‘한국 화장품’ 검색 트래픽 정보를 각각 데이터 원천 페이지로부터 가져와서 지금까지의 시계열을 출력한 다음에 사용자의 요청에 따라 예측치를 동적으로 표시해준다. ‘AIRPORT FORECASTING’은 전국 공항에 입국하는 유커 비율 시각화 서비스의 경우 공식적으로 발표되는 수치를 활용하여 지도에 알기 쉽게 원 그래프로 표시한다. ‘TRAFFIC FORECASTING’은 바이두 트래픽이 높은 상위 콘텐츠 몇 가지를 엄선하여 보여준다. ‘MIND MAP’은 연관검색어 클라우드이다. ‘한국여행’이라는 키워드에서부터 시작하여 바이두의 연관검색어 알고리즘에 따라서 중국 유저들이 많이 찾는 연관검색어를 마인드맵 형태로 보

여준다. 클릭하여 동적으로 심도를 높여가며 볼 수 있는 것이 특징이다.

‘VARIABLE FORECASTING’은 관광수요 시계열 그래프 위에 수요 예측 값을 출력하게 된다. <Figure 4>가 해당 서비스를 사용할 때의 화면이다. 두 개의 그래프 창이 있고 왼쪽 그래프 창은 독립변수 측정치의 추이를 보여주고 오른쪽은 중국인 관광수요의 추이를 보여준다. js슬라이더를 탑재하여 사용자가 독립변수를 임의로 조절하여 값을 제출하면 입력한 변수 값이 파싱되어 앞서 도출한 중국인 입국자 수 회귀모델에 대입하여 모델에 고정된 가중치가 곱해진다. 계산된 수요 예측치는 동적으로 오른쪽 그래프에 가장 오른쪽 값으로 표시된다. 이 기능을 통해 유저는 독립변수에 따라서 유동적으로 달라지는 미래 수요 예상치를 확인할 수 있다. 관광수요 예측뿐만 아니라 정책결정자가 독립변수를 조절



〈Figure 3〉 Index Page of ‘Youke Mining’



〈Figure 4〉 VARIABLE FORECASTING: Regression Graph of Prediction Model

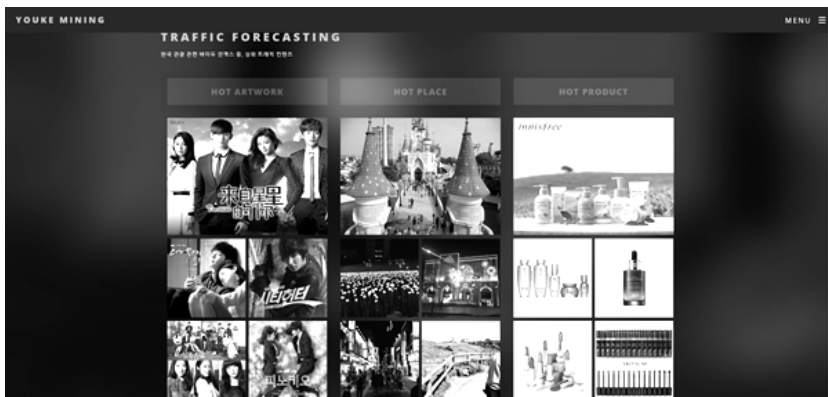
해가며 관광수요를 증가시킬 수 있는 결정을 할 때 도움이 될 수 있다.

〈Figure 5〉의 ‘AIRPORT FORECASTING’은 공식적으로 발표된 공항별 중국인 유커의 입국자 수 비중을 활용한다. 해당 공항이 있는 곳의 지리적 위치에 파이 차트의 크기에 따라 입국자 수를 시각화하여 나타내기 때문에 직관적으로 알기 쉽게 표현된다. 데이터에 의해 그려진 이 서비스의 지도로 지방정부의 입안자가 각 지역의 유커 입국 현황을 확인할 수 있다.

〈Figure 6〉에서 볼 수 있는 ‘TRAFFIC FORECASTING’ 서비스는 한국 드라마, 한국 관광 명소, 한국 화장품 세 가지 카테고리에서 바이두 웹검색 트래픽 정보에서 데이터를 기준으로 트래픽이 가장 높은 콘텐츠를 보여준다. 각 썸네일 이미지를 클릭하면 실제 해당 키워드로 검색한 결과 창이 중국 바이두 사용자가 많이 검색해본 키워드들이 표시되므로 중국 내 한류의 현황이 어떠한지, 관심 있는 여행지는 어느 곳인지, 중국인들이 많이 찾는 한국 화장품 브랜드는 어



〈Figure 5〉 AIRPORT FORECASTING: Local Youke Rate of All the Airport



〈Figure 6〉 TRAFFIC FORECASTING: Top Topics based on Baidu Search

편 것인지 파악할 수 있다.

한국화장품은 본 연구에서 세운 모형 식에도 ‘한국화장품’이라는 키워드의 웹검색 트래픽 정보가 수요 예측치를 계산하는 데에 독립변수로 사용된다. 바이두에서 한국관광을 가장 많이 검색한 군집으로 밝혀진 30대 여성들의 관광 양상을 보면 쇼핑이 큰 비중을 차지했다. 유커의 국내 쇼핑은 국내 산업에 큰 부가가치를 안겨주기 때문에 관심 있게 살펴볼 필요가 있다.

‘MIND MAP’은 바이두 웹검색결과를 기반으

로 연관검색어를 마인드맵 형태로 재구성한 것이다. 연관검색어는 유저가 해당 키워드와 함께 검색한 키워드 정보를 저장하여 알고리즘에 따라 연관 있는 검색어로 추천하는 방식이다. <Figure 7>은 한국 관광과 직접적인 관련이 있는 ‘한국 여행’ 키워드에 대한 연관검색어가 나뭇가지처럼 달려있다. 각 키워드를 클릭하면, 클릭한 키워드의 연관검색어들이 표현되면서 동적으로 변화하는 화면이다. ‘한국 여행’에서 시작하여 그에 따른 연관검색어들을 연쇄적으로 탐색할



(Figure 7) MIND MAP: Graphic of Related Keyword from Baidu Search

수 있으며 실제 중국인들이 어떤 키워드와 함께 검색하며 관심을 가졌는지 알 수 있다.

정책결정자가 쉽게 활용할 수 있는 솔루션 시스템 ‘유커마이닝’은 중국인 관광객 유치 향상을 위해 유용한 서비스를 제공하는 Thin-client 솔루션으로서 간단한 클릭과 빠르고 쉬운 인터페이스로 의사결정에 도움을 줄 수 있을 것이다.

## 5. 결론

최근 정보기술의 발전에 발맞추어 빠르게 진화하고 있는 데이터 수집/가공/분석의 기술들은 소비자의 의사결정, 기업의 의사결정, 특정 이벤트에 대한 예측, 적발 등 다양한 분야에서 다채로운 형태로 데이터기반의 의미있는 인사이트를 도출하는 데 기여하고 있다(Seo and Park, 2017; Jun et al. 2017). 또한 이러한 데이터 기반의 인사이트들은 기존의 전통적인 방법인 설문이나 실험을 통하여 조직행동(Kwahk and Park, 2016), 소비자 구매(Park and Park, 2013) 및 정보탐색

활동(Lee et al. 2014), 온라인 구전활동(Park and Kim, 2008; Park and Lee, 2008; Park et al. 2007) 등에서 도출된 결과를 재검증해 주거나 기존에 도출하지 않았던 여러 현상들도 찾아내 준다. 이에 맞추어 본 연구는 웹검색 트래픽 정보의 활용 가능성을 국내의 중국인 관광수요 예측 방면에 확장하고 실용적으로 도입하였다. 구체적으로, 본 연구는 웹검색 트래픽 정보를 중국인 관광수요 예측에 이용하여 중국인 관광 수요 예측 식을 개발하고, 웹검색 트래픽 정보가 포함된 예측식이 상대적으로 뛰어난 성능을 보인다는 가설을 검증하기 위해 기존 연구들에서 유의하다고 검증된 변수들을 이용한 모델과 본 연구에서 검증하고 하는 트래픽 변수들을 이용한 모델들의 예측력과 유의확률을 비교를 보여주고 있다.

본 연구는 이론적, 실무적, 정책적인 세 가지 관점에서 의미 있는 시사점을 가지고 있다. 이론적인 관점에서, 본 연구는 중국인의 웹검색 트래픽 정보를 기반으로 하여 인바운드 관광수요를 설명한 첫 시도라는 점이다. 지금까지 연구되어 온 예측모형에는 주로 자기 시계열을 활용한 연

구, 외부 거시적인 변수들을 사용한 연구가 대부분인데, 본 연구에서는 입국자 통계를 기존 연구에서 제안하는 변수에 유커 여행을 설명할 수 있는 검색 키워드의 웹검색 트래픽 정보를 추가하여 예측력을 높였다는 차원에서 이론적인 공헌이 있다. 실무적인 관점과 정책적인 관점에서 의미 있는 시사점을 가진다. 실무적인 측면에서, 본 연구는 중국 최대 점유율의 검색엔진 바이두에서 제공하는 검색 트래픽 정보를 활용하여 우리나라가 중국인의 관광수요를 예측하기 위한 방법을 제안했다. 중국 웹검색 엔진 시장의 80%인 바이두의 검색 트래픽을 이용하는 것은 많은 중국인들이 우리나라에 방문하는 이유를 실시간으로 가장 잘 반영할 수 있다는 것을 나타내고 있다. 또한 쉽게 수집 가능한 웹검색 트래픽 정보와 환율 정보만으로도 상당부분 중국인 입국자를 예측할 수 있다고 제시하는 본 연구의 유커 수요 예측 식은 실무적인 가치를 더욱 더 증가시켜준다. 정책적인 측면에서, 본 연구는 늘어나는 중국인 관광객인 유커의 수요를 예측함으로써 관광정책에 이용할 수 있다. 관광 정책이란 관광 자원을 바탕으로 관광객들을 만족시키기 위해 관광 서비스를 제공하려는 시책이다. 중국인 관광객의 수요가 예측이 된다면 중국인들의 관광 욕구를 충족하기 위해 어떠한 정책을 시행해야 할지 정할 때 이용하고 자원의 효율적인 관리로 수익 극대화의 기회 및 성공가능성 향상을 위한 의사결정에 도움을 줄 수 있을 것이다.

본 연구는 다음과 같은 한계점을 지니고 있다. 첫째, 바이두에서 제공받는 웹검색 트래픽 정보는 주 단위로 획득할 수 있지만, 중국인 입국자 통계는 월 단위로 획득되어, 실시간성을 반영할 수 있는 주간 예측을 수행할 수 없다는 점이다. 주간 입국자 통계가 제공된다면, 좀 더 실시간에

가까운 동적인 예측 및 인사이트 도출이 가능할 것이다. 둘째, 입국자 통계의 경우 관광 목적이 아닌 다른 사람들이 포함될 수 있는 점이다. 입국자 중 여행 목적인 사람들을 구분하여 데이터를 축적할 수 있으면, 본 연구에서 제안하는 모형의 설명력이 증진될 것으로 예상된다. 마지막으로, 본 연구에서 검색키워드 선정 시 모든 키워드를 고려하지 못하고, 인터뷰를 통해 일부 키워드만 고려했다는 점이다. 이런 키워드 선정의 한계는 웹검색 트래픽 정보 연구에서 항상 제기되는 한계점으로(Jun and Park, 2017), 정성적인 인터뷰 및 연구자의 직관을 통해 해결될 수밖에 없으므로, 최대한 가능한 키워드 풀 안에서 시행착오를 통해 도출할 수밖에 없다고 할 수 있다. Jun and Park (2013)의 지침에 따라, 본 연구에서도 정성적인 인터뷰, 다수의 후보군 도출 및 시행착오 통한 키워드 선정 과정을 빠짐없이 수행하였기 때문에, 조금이나마 한계를 극복할 수 있었다.

## 참고문헌(References)

- Ahn. Y., D.-H. Sohn. Y.-T. Lee and C.-K. Lee, "The Economic Impact of Tourism Industry in Korea - Input - Output Analysis," Korean Public Administration Review, Vol. 29, No. 1(1995), 123-142.
- Archer, B. Sustainable Tourism: An Economist's Viewpoint, Sustainable Tourism in Islands and Small States, Pinter, London, 1996.
- Bae. S.-K., S.-W. Kim and H.-J. Kim, Contracting-out Recreational Facilities Service Management, Korean Contracting-out Management Institute, Seoul, 2015.



- Choi, H. and H. Varian, "Predicting the Present with Google Trends," *Economic Record*, Vol. 88, No. 1(2012), 2-9.
- Dharmaratne, G. S. "Forecasting Tourist Arrivals in Barbados," *Annals of Tourism Research*, Vol. 22, No. 4(1995), 804-818.
- Ginsberg, J., M.H. Mohebbi, R.S. Patel, L. Brammer, M.S. Smolinski and L. Brilliant, "Detecting Influenza Epidemics Using Search Engine Query Data," *Nature*, Vol. 457, No. 7232(2009), 1012-1014.
- Jang, B.-K. "China Outbound Market Trend and Strategies for Tour Competitiveness Enhancement," *Korea Tourism Policy*, Vol. 62(2015), 10-21.
- Jeong E.-S., T. Liu. and S.-H. Cho, "Demand Forecasting of Inbound Tourism Revenues in China," *Journal of Hotel Resort*, Vol. 12, No. 1(2013), 135-157.
- Jun, S.-P., D. Choi, H.-W. Park, B.-G. Seo and D.-H. Park, "Development of Systematic Process for Estimating Commercialization Duration and Cost of R&D Performance," *Journal of Intelligence and Information Systems*, Vol. 23, No. 2(2017), 139-160.
- Jun, S.-P. and D.-H. Park, "Visualization of Brand Positioning Based on Consumer Web Search Information: Using Social Network Analysis," *Internet Research*, Vol. 27, No. 2(2017), 381-407.
- Jun, S.-P. and D.-H. Park, "Consumer Information Search Behavior and Purchasing Decisions: Empirical Evidence from Korea," *Technological Forecasting and Social Change*, Vol. 107(2016), 97-111.
- Jun, S.-P., D.-H. Park and J. Yeom, "The Possibility of Using Search Traffic Information to Explore Consumer Product Attitudes and Forecast Consumer Preference," *Technological Forecasting and Social Change*, Vol. 86(2014), 237-253.
- Jun, S.-P. and D.-H. Park, "Intelligent Brand Positioning Visualization System Based on Web Search Traffic Information : Focusing on Tablet PC," *Journal of Intelligence and Information Systems*, Vol. 19, No. 3(2013), 93-111.
- Kang, T. and D.-H. Park, "The Effect of Expert Reviews on Consumer Product Evaluations : A Text Mining Approach," *Journal of Intelligence and Information Systems*, Vol. 22, No. 1(2016), 63-82.
- Kim, B., Y. Choi and D.-H. Park, "Investment Model Development Based on Web-search Traffic Information: Focusing on KOSPI Index," *Entrue Journal of Information Technology*, Vol. 14, No. 3(2015), 63-81.
- Koo, P. and M. Kim, "A Study on the Relationship between Internet Search Trends and Company's Stock Price and Trading Volume," *The Journal of Society for e-Business Studies*, Vol. 20, No. 2(2015), 1-14.
- Kwahk, K.-Y. and D.-H. Park, "The Effects of Network Sharing on Knowledge-sharing Activities and Job Performance in Enterprise Social Media Environments," *Computers in Human Behavior*, Vol. 55(2016), 826-839.
- Lazer, D., R. Kennedy, G. King and A. Vespignani, "The Parable of Google Flu: Traps in Big Data Analysis," *Science*, Vol. 343, No. 6176(2014), 1203-1205.
- Lee C.-K., K.-S. Song and H.-J. Song, "Determinants of Bi-national Tourism

- Demand from Japan to Korea: Using Econometric Models,” *Journal of Tourism & Leisure Research*, Vol. 18, No. 4(2006), 7-25.
- Lee, D., T. Kang and D.-H. Park, "The Research on PC-based Versus Mobile Device-based Shopping Behavior Depending on Consumer Purchase Decision Process: Focusing on Task-Technology Fit Theory," *Entrue Journal of Information Technology*, Vol. 13, No. 3(2014), 107-122.
- Lee D.-S. and S.-J. Jung, “An Analysis on Appropriateness of Performance Indicators in the Certification System for Good Travel Program,” *Proceeding of 55<sup>th</sup> International Tourism Conference(2004)*, 491-500.
- Lui, C., P.T. Metaxas and E. Mustafaraj, “On the Predictability of the US Elections through Search Volume Activity,” *Proceedings of the IADIS International Conference on e-Society(2011)*.
- Park. D.-H. and M.-S. Lee, “Development of Forecasting Model for Travel Demand Using Web Search Traffic Information: Focusing on Major Cities in Gangwon-do,” *Gang-won Economy Brief*, Vol. 2(2015), 515-542.
- Park, D.-H. and S. Kim, “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), 399-410.
- Park, D.-H. and J. Lee, “eWOM Overload and its Effect on Consumer Behavioral Intention Depending on Consumer Involvement,” *Electronic Commerce Research and Applications*, Vol. 7, No. 4(2008), 386-398.
- Park, D.-H., J. Lee and I. Han, “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), 125-148.
- Park, S.-B. and D.H. Park, “The Effect of Low-versus High-variance in Product Reviews on Product Evaluation,” *Psychology & Marketing*, Vol. 30, No. 7(2013), 543-554.
- Seo, B.-G. and D.-H. Park, "Development on Early Warning System about Technology Leakage of Small and Medium Enterprises," *Journal of Intelligence and Information Systems*, Vol. 23, No. 1(2017), 143-159.
- Sheldon, P.J. and T. Var, “Tourism Forecasting: A Review of Empirical Research,” *Journal of Forecasting*, Vol. 4, No. 2(1985), 183-195.
- Song. J.H. and H.J. Huh, “Destination Repositioning of Jeju Island Resort: Comparison with Other Domestic Competitive Ones,” *Journal of Tourism Sciences*, Vol. 27, No. 3(2003), 9-24.
- Uysal, M. and J.L. Crompton, “An Overview of Approaches Used to Forecast Tourism Demand,” *Journal of Travel Research*, Vol. 23, No. 4(1985), 7-15.
- Vosen, S. and T. Schmidt, “Forecasting Private Consumption: Survey-Based Indicators vs. Google Trends,” *Journal of Forecasting*, Vol. 3, No. 6(2011), 565-578.
- Witt, S.F. and C.A. Witt, *Modeling and Forecasting Demand in Tourism*, Academic Press Ltd, London, 1992.

## Abstract

# Development of Yóukè Mining System with Yóukè's Travel Demand and Insight Based on Web Search Traffic Information

Choi, Youji\* · Park, Do-Hyung\*\*

As social data become into the spotlight, mainstream web search engines provide data indicate how many people searched specific keyword: Web Search Traffic data. Web search traffic information is collection of each crowd that search for specific keyword. In a various area, web search traffic can be used as one of useful variables that represent the attention of common users on specific interests. A lot of studies uses web search traffic data to nowcast or forecast social phenomenon such as epidemic prediction, consumer pattern analysis, product life cycle, financial invest modeling and so on. Also web search traffic data have begun to be applied to predict tourist inbound. Proper demand prediction is needed because tourism is high value-added industry as increasing employment and foreign exchange. Among those tourists, especially Chinese tourists: Youke is continuously growing nowadays, Youke has been largest tourist inbound of Korea tourism for many years and tourism profits per one Youke as well. It is important that research into proper demand prediction approaches of Youke in both public and private sector. Accurate tourism demands prediction is important to efficient decision making in a limited resource. This study suggests improved model that reflects latest issue of society by presented the attention from group of individual. Trip abroad is generally high-involvement activity so that potential tourists likely deep into searching for information about their own trip. Web search traffic data presents tourists' attention in the process of preparation their journey instantaneous and dynamic way. So that this study attempted select key words that potential Chinese tourists likely searched out internet. Baidu-Chinese biggest web search engine that share over 80%- provides users with accessing to web search traffic data. Qualitative interview with potential tourists helps us to understand the information search behavior before a trip and identify the keywords for this study. Selected key words of web search traffic are categorized by how much directly related to "Korean Tourism" in a three levels. Classifying categories helps to find out which keyword can

---

\* Graduate School of Business IT, Kookmin University

\*\* Corresponding Author: Park, Do-Hyung

College of Business Administration/ Graduate School of Business IT, Kookmin University

Tel: +82-2-910-5613, Fax: +82-2-910-4560, e-mail: dohyungpark@kookmin.ac.kr

explain Youke inbound demands from close one to far one as distance of category. Web search traffic data of each key words gathered by web crawler developed to crawling web search data onto Baidu Index. Using automatically gathered variable data, linear model is designed by multiple regression analysis for suitable for operational application of decision and policy making because of easiness to explanation about variables' effective relationship. After regression linear models have composed, comparing with model composed traditional variables and model additional input web search traffic data variables to traditional model has conducted by significance and R squared. after comparing performance of models, final model is composed. Final regression model has improved explanation and advantage of real-time immediacy and convenience than traditional model. Furthermore, this study demonstrates system intuitively visualized to general use -Youke Mining solution has several functions of tourist decision making including embed final regression model. Youke Mining solution has algorithm based on data science and well-designed simple interface. In the end this research suggests three significant meanings on theoretical, practical and political aspects. Theoretically, Youke Mining system and the model in this research are the first step on the Youke inbound prediction using interactive and instant variable: web search traffic information represents tourists' attention while prepare their trip. Baidu web search traffic data has more than 80% of web search engine market. Practically, Baidu data could represent attention of the potential tourists who prepare their own tour as real-time. Finally, in political way, designed Chinese tourist demands prediction model based on web search traffic can be used to tourism decision making for efficient managing of resource and optimizing opportunity for successful policy.

**Key Words** : Chinese, Youke, Tourist Demand, Web Search Traffic, Baidu, Bigdata, Inbound

Received : September 1, 2016 Revised : June 28, 2017 Accepted : June 28, 2017

Publication Type : Regular Paper Corresponding Author : Park, Do-Hyung

## 저 자 소개



### 최유지

현재 국민대학교 Business IT 전문대학원에서 Customer Experience트랙으로 석사과정에 재학 중이다. 주요 관심 분야는 컨텍스트에 따라서 최적의 방법론을 적용하는 데이터 기반 사용자 분석(Customer Analytics), 사용자 분석 시스템 설계, 사용자 경험 디자인(User eXperience) 등이다. 개인의 관심이 집단적으로 작용하는 방식과 그 정도를 정량화하는 방법에 대해 연구한다.



### 박도형

KAIST 경영대학원에서 MIS 전공으로 석사/ 박사학위를 취득하였다. 현재 국민대학교 경영대학 경영정보학부/ 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 담당했었고, 스마트폰, 스마트TV, 스마트Car 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.