

SNS 데이터를 활용한 관광지 혼잡도 및 방문자 특성 분석에 관한 연구⁺

(A Study on the Analysis of the Congestion Level of Tourist
Sites and Visitors Characteristics Using SNS Data)

이 상 훈¹⁾, 김 수 연^{2)*}
(Sang Hoon Lee and Su-Yeon Kim)

요 약 SNS는 일상생활에 매우 밀접한 서비스가 되었다. SNS를 통해 마케팅이 이루어지면서 흔히 핫플레이스라 불리는 장소가 생겨나고, 이곳으로 사용자들이 몰리고 있다. 하지만 단기간 많은 사람이 몰리며 혼잡한 경우가 빈번하게 발생하여 방문자와 서비스 제공자 모두 부정적인 경험을 하게 되는 경우가 많다. 이러한 문제를 개선하기 위해 혼잡도를 파악해야 하지만 개인적 수준에서 특정 지역의 혼잡도를 알아볼 방법은 매우 한정적이다. 이에 본 연구에서는 SNS상의 데이터를 활용하여 특정 관광지에 대한 혼잡도 정보 및 방문자들의 특성을 파악할 수 있는 시스템을 제시하고자 하였다. 이를 위해 사용자가 업로드한 포스팅 데이터와 이미지 분석을 사용하였으며 네이버 DataLab 시스템을 이용하여 제안 시스템의 성능을 검증하였다. 관광지 유형별로 3개 장소를 선정하여 비교 검증한 결과 본 연구에서 산출한 결과와 DataLab에서 제공하는 혼잡도 수준이 유사한 것으로 나타났으며, 특히 본 연구는 특정 기업이나 서비스에 종속되지 않는 사용자의 실 데이터에 기반한 혼잡도를 제공하였다는 것에 의의가 있다.

핵심주제어: 소셜네트워크서비스(SNS), 혼잡도, 포스팅 데이터 분석, 이미지 분석, 방문자 특성

Abstract SNS has become a very close service to our daily life. As marketing is done through SNS, places often called hot places are created, and users are flocking to these places. However, it is often crowded with a large number of people in a short period of time, resulting in a negative experience for both visitors and service providers. In order to improve this problem, it is necessary to recognize the congestion level, but the method to determine the congestion level in a specific area at an individual level is very limited. Therefore, in this study, we tried to propose a system that can identify the congestion level information and the characteristics of visitors to a specific tourist destination by using the data on the SNS. For this purpose, posting data uploaded by users and image analysis were used, and the performance of the proposed system was verified using the Naver DataLab system. As a result of comparative verification by selecting three places by type of tourist destination, the results calculated in this study and the congestion level provided by DataLab were found to be similar. In particular, this study is meaningful in that it provides a degree of congestion based on real data of users that is not dependent on a specific company or service.

Keywords: Social Network Service(SNS), Congestion Level, Posting Data Analysis, Image Analysis, Visitor Characteristics

* Corresponding Author: sykim@daegu.ac.kr

+ 이 논문은 2019학년도 대구대학교 학술연구비지원에 의한 논문임.
Manuscript received September 23, 2022 / revised October

20, 2022 / accepted October 22, 2022

1) 대구대학교 컴퓨터정보공학부, 제1저자

2) 대구대학교 컴퓨터정보공학부, 교신저자

1. 서론

과거에는 특정 관광지에 대한 정보를 수집하기 위해서는 검색엔진을 활용한 방법이 주로 사용되었다. 하지만 SNS의 보급으로 인해 젊은 층을 중심으로 관광지에 대한 정보 검색에 SNS를 활용하고 있다. 이들은 SNS상에 업로드되는 짧은 글과 이미지로 관광지에 대한 정보를 확인하고 관광지 방문 여부를 선택한다. SNS 이용자들은 검색엔진이나 관광지 추천 서비스보다 SNS상의 게시글을 활용하여 방문하고자 하는 곳을 정하고 있다. 소규모 매장들은 이러한 점을 인지하고 활발하게 SNS를 통한 마케팅 활동을 펼치고 있으며, 지자체 또한 다양한 SNS 채널을 통해 관광지를 홍보하고 있다(Kim et al., 2021; Herrero, 2018; Kim and Jang, 2021). 하지만 막상 관광지에 방문했을 때 너무 많은 방문객으로 인해 정상적인 관광이 불가능한 경우가 증가하고 있다. 다수 이용자들이 SNS를 통하여 가장 인기 있는 관광지를 선택하고 그곳으로 방문하기 때문이다. 이런 경우 관광지에 대한 부정적인 인상이 생기기 쉽고, 방문객들 또한 혼잡함으로 인해 불편을 겪게 된다(Saenz-de-Miera, 2012).

본 연구에서는 특정 관광지의 혼잡도 수준을 파악하기 위하여 SNS상에 업로드되는 정보를 활용하여 이용자들에게 예상되는 혼잡도를 제공하고자 한다. 또한 SNS상의 데이터를 활용하여 관광지에 방문하는 방문객의 특성을 제공함으로써 관광지의 분위기를 알려주고자 한다. 본 논문의 구성은 다음과 같다. 2장에서 이론적 배경을 살펴보고 3장에서는 연구모형을 설계하고 구현하였다. 4장에서 시스템의 검증 실시하였고, 마지막으로 결론 및 향후 연구를 제시한다.

2. 이론적 배경

2.1 SNS

SNS는 Social Networking Service의 약자로, Social Media의 분류 중 하나이다. SNS는 이제 우리들의 삶에서 필수적인 요소로 자리 잡았으며, 10대에서부터 50대 이상까지 폭넓은 연령층이 다양한 서비

스를 활용하고 있다. SNS는 자신의 의견을 전파하고자 할 때 매우 효과적인 도구로 사용되고 있으며(Jeon and Kwahk, 2016; Lee et al., 2016) 수많은 연예인이나 정치인들도 활용하고 있다. 유명인이나 사회 주요 인사가 아니더라도 다양한 개인들은 온라인 상에서 자신을 드러내기 위하여 자신만의 포스팅을 업로드하거나 라이프스타일에 대한 동영상과 이미지 등을 업로드하여 온라인상의 새로운 자아를 만들어가고 있다(Song and Shin, 2022; Lee and Kwahk, 2021; Lee, 2018). 실제로 카페나 식당뿐 아니라 유적지와 박물관, 공공기관에 이르기까지 인증샷의 용도로 사용되고 있다(Kim, 2021). 온라인 활동이 늘어난 최근에는 SNS의 영향력이 점차 더 커지고 있다(Kim et al., 2021; Hwang et al., 2020). SNS의 영향력이 지속적으로 증가하면서 기업의 마케팅 수단으로 활용되고 있다. 대형 규모의 브랜드뿐 아니라 카페나 헤어, 네일 메이크업 등의 소규모 매장들도 SNS상에서 홍보를 하고 있으며, 다수 연구에서 실제로 SNS상의 광고가 효과가 있음을 확인할 수 있다(Lee et al., 2022; Oh, 2022; Lee and Mo, 2022). 최근에는 지자체와 공공기관들도 SNS상에서 고객들과 소통하고 있고(Kwak, 2021; Kim and Oh, 2021; Yang and Kang, 2022), 여러 관광지들도 SNS를 통해 마케팅을 펼치고 있으며 긍정적인 효과를 내고 있다(Kwak, 2021; Jun and Kim, 2021; Kim and Lee, 2021; Kim and Choi, 2021)

SNS에 업로드되는 데이터가 많아지면서, 기존에 검색엔진을 통해 특정 방문지에 대한 정보를 수집하던 패러다임이 젊은 층을 중심으로 SNS상에서 관련 정보를 얻는 형태로 변화하고 있다(Kim and Park, 2021; Kim and Lee, 2021; Kang and Chon, 2021). 검색엔진을 통해 정보를 수집하는 것이 아니라 본인에게 익숙한 플랫폼에서 구매하고자 하는 서비스 또는 제품, 방문하고자 하는 곳에 대한 정보를 확인하고 있으며, 쇼핑물이 아니라 SNS를 통한 상품구매도 증가하고 있다(Son and Heo, 2018). 특정 지역을 방문할 경우 ‘핫플’이라 불리는 인기 장소를 찾기 위해 SNS를 검색하고 찾아가는 사용자들도 늘고 있다(Kim et al., 2021; Herrero, 2018; Kim and Jang, 2021). 하지만 SNS를 통해 확인한 관광지에 방문했을 때 너무 많은 방문객으로 인해 정상적인 관광이 불가능한 경우가 점차 늘어나고 있다. 다수 사용자들

이 SNS를 통하여 가장 핫한 관광지를 선택하고 거기로 몰리는 현상이 발생하기 때문이다. 이 경우 관광지에 대한 부정적인 인상이 생기기 쉽고 방문객들 또한 불편을 겪게 된다(Saenz-de-Miera, 2012). 이에 본 연구에서는 SNS에 업로드되는 데이터를 활용하여 사용자가 미리 방문하고자 하는 관광지의 혼잡도를 예측할 수 있도록 하여 관광지에 대한 긍정적 경험을 제고하고자 한다.

2.2 관광지 혼잡도

혼잡도란 일반적으로 특정 공간에 정해진 수용인원 대비 방문 또는 서비스 사용 중인 인원의 비율을 의미한다. 심리학적으로는 특정 공간 내에 사람이나 물건 등으로 인해 개인의 행동에 제약이 받거나 부정적으로 인식하는 경우로 정의하고 있다(Stokols, 1987). 혼잡은 인적 혼잡과 공간적 혼잡으로 나눌 수 있는데, 인적 혼잡은 많은 사람으로 인해 혼잡한 것을 뜻하며 공간적 혼잡은 공간 내 다양한 오브젝트로 인해 제약받는 것을 의미한다(Hui and Bateson, 1991; Machleit et al., 1994; Li et al., 2009).

관광지 혼잡도란 관광지에 방문 중인 사람으로 인해 혼잡한 정도를 의미하며, 일반적으로 관광지 내 혼잡도가 높아질수록 부정적 경험이 증가하고 혼잡지역을 이탈하려는 경향이 있다(Manning and Ciali, 1980; Saenz-de-Miera, 2012; Yang et al., 2019). 또한 관광지나 카페, 레스토랑 등에서 혼잡도가 높을 경우 만족도에 부정적인 영향을 미치게 되고 구매나 재방문 의도에도 영향을 미친다는 것을 밝혀낸 연구도 있다. Han(2020)은 외식 프랜차이즈 매장에서의 혼잡도 연구를 수행하였는데 점포가 혼잡할수록 고객들의 지각된 기대 가치가 낮아지고 이는 결국 재방문 의도를 떨어뜨린다는 것을 밝혀냈다. Quan and Han(2019)의 연구에서도 인적, 공간적 혼잡도 모두 고객 만족도에 부정적인 영향을 미친다는 것을 밝혔으며, 제주도에서 위치한 우도의 방문자를 대상으로 한 Yang et al.(2019)의 연구에서는 혼잡도가 만족도에 영향을 주는 것으로 나타났다. 특히 최근 SNS의 영향력에 힘입어 핫플레이스 위주로 방문객들이 증가하여 짧은 순간 엄청난 인파가 몰려 교통이 마비되고 서비스 대기 시간이 늘어남 등 부정적인 현상이 증가하고 있다.

이처럼 음식점, 카페, 박물관 등과 같은 관광지의 경우 혼잡도가 사용자 만족에 부정적인 영향을 주고 있으나 개인이 특정 위치의 혼잡도를 사전에 파악하는 것은 매우 어렵다. 방문자들에게 혼잡도 정보를 제공하기 위하여 리뷰 데이터, 검색 데이터 등을 활용한 연구들이 수행되었다. Hu et al.(2022)의 연구는 여행자들이 온라인상에 게시한 리뷰를 기반으로 혼잡도를 예측하고자 하였다. 이 연구에서는 홍콩을 방문하는 여행자들을 대상으로 호텔 이용 후기와 마켓 이용 후기를 추출한 다음, 주간/월별로 게시된 후기 수와 별점을 활용하여 혼잡도를 예측하고자 하였다. Yang et al.(2021)은 중국의 Baidu 검색 데이터를 활용하여 관광 수요와 관련성이 높은 키워드를 세그먼트별로 분석함으로써 혼잡도를 예측하고자 하였으며, Li et al.(2017)의 연구는 베이징을 방문하는 관광객의 수를 예측하기 위하여 검색엔진을 활용하였다. Xie et al.(2021)은 중국 크루즈 관광 수요를 예측하기 위한 빅데이터 기반의 혼잡도 예측 모델을 제시하였다.

이상에서 살펴본 바와 같이 혼잡도를 파악하기 위한 연구가 진행되고 있고 구글과 네이버 같은 포털 사이트 역시 자사의 검색엔진 키워드 데이터를 활용하여 부분적인 혼잡도 정보를 제공하고 있다. 하지만 위에서 제시된 방법은 SNS를 통해 단기간에 인기가 높아지는 핫플레이스에 대한 혼잡도를 빠르게 반영할 수 없다는 단점이 있다. 이에 본 연구에서는 SNS상의 데이터를 활용하여 시시각각 변하는 혼잡도 정보를 사용자에게 신속하게 제공하고자 한다.

3. 시스템 설계 및 구현

3.1 연구 방법

본 연구에서는 SNS 데이터를 활용하여 방문자들에게 관광지의 혼잡도와 방문객 특성을 제공하기 위하여 Fig. 1과 같은 단계를 거쳐 시스템을 설계하였다.

먼저 특정 관광지에 대한 SNS상의 데이터를 수집한다. 이때 수집되는 데이터는 업로드된 이미지와 업로드 시간, 업로드 위치 등의 포스팅 데이터이다. 이후 수집된 이미지와 게시글 데이터를 추출한다. 수집된 데이터는 게시글 데이터에 기반한 혼잡도 분석,

이미지를 통한 혼잡도 분석, 이미지를 통한 방문자 특성 분석단계를 거치게 된다. 각 분석단계에서 도출된 결과값을 활용하여 관광지에 대한 최종 혼잡도와 방문자 특성 정보를 이용자에게 제공한다.

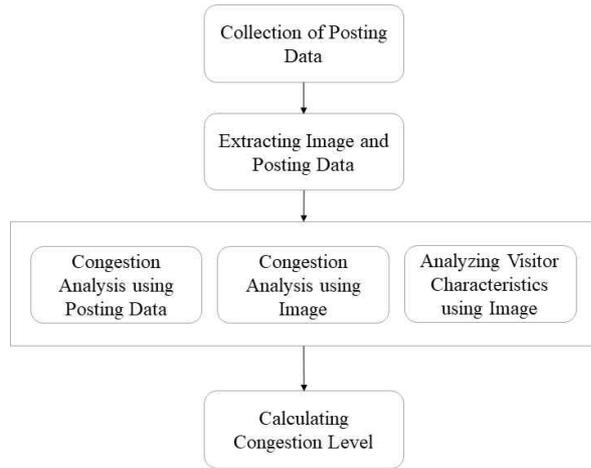


Fig. 1 Research Procedure

3.2 관광지의 포스팅 데이터 수집

특정 관광지에 대한 혼잡도를 분석하기 위해 가장 먼저 SNS상에 업로드된 포스팅 데이터를 수집해야 한다. 본 연구에서는 국내에서 가장 많은 사용자를 보유한 이미지 기반 SNS 서비스인 인스타그램을 통해 데이터를 수집하였다. 관광지 위치에 대한 정확한 포스팅 데이터를 수집하기 위하여 사용자가 직접 특정 위치에 있다고 위치정보를 설정한 포스팅을 대상으로 수집을 진행하였다. 사용자가 특정 위치에 있다고 직접 지정한 경우 일반적인 해시태그 기반의 위치정보에 비해 특정 지역에 대한 보다 정확한 정보를 획득할 수 있어 사용자가 직접 설정한 경우의 포스팅만 수집하였다.

인스타그램의 경우 다양한 API가 준비되어 있으나 사업자를 보유하고 있으며 자체 심사를 통과한 업체에 대해서만 API를 사용할 수 있도록 하고 있다. 본 연구에서는 API 대신 개인계정을 활용하여 데이터를 수집하고 이를 토대로 연구를 진행하였다. 데이터 수집은 파이썬과 Selenium을 이용하였고 제주도내 위치한 유명 카페의 포스팅 데이터를 활용하였다. 수집된 데이터는 총 3,138개의 이미지와 14개월간(2021.6~2022.8.)의 포스팅 데이터이다. 수집된 포스팅 데이

터는 user_ID, posintg_content, hashtag, uploaded_date, reply, like_count, image file 등을 포함한다.

3.3 포스팅 내 이미지와 게시물 데이터 추출

앞서 수집한 데이터 중에서 개인을 특정할 수 있는 계정명과 같은 데이터와 불필요한 내용을 제외하고 Fig. 2와 같이 이미지, 업로드 위치, 게시날짜 데이터만 추출하여 연구에 활용하였다.

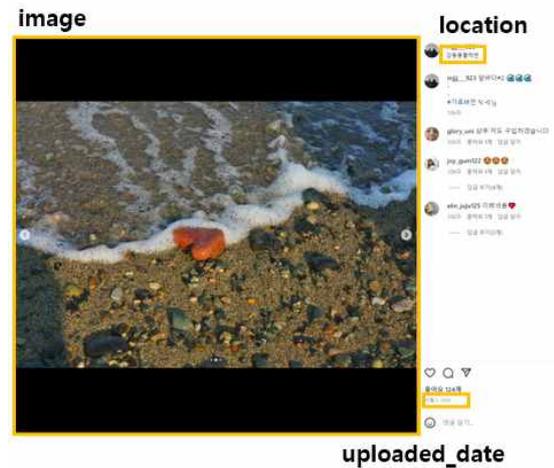


Fig. 2 Data Collected from SNS

이미지의 경우 기존에 인스타그램에서 자체적으로 분류된 파일명으로 설정된 것을 연구 진행을 위하여 일괄적으로 변경하였으며 추출된 포스팅 데이터와 연결될 수 있도록 하였다. 또한 추출된 포스팅 데이터의 경우 텍스트 형식으로 구성되어 있으나 이를 연구에 활용하기 위하여 시계열 형태의 데이터로 변환하였다.

3.4 게시물 데이터 분석을 통한 혼잡도 분석

게시글 데이터 분석을 위해 본 연구에서는 파이썬과 Pandas, Numpy를 이용하였다. 수집된 데이터의 경우 동일한 날짜에 여러 개의 포스팅이 있을 수 있고 특정 날짜에는 하나도 없을 수 있기에 Pandas의 데이터프레임 형태로 변경하고 게시글의 업로드 시간을 기준으로 일간, 월간, 연간 데이터 등으로 분류하였다. 분류 결과는 Fig. 3과 같다.

0	2021-06-08	3	Friday	388	2021-06-13	57	2021-06-30	275
1	2021-06-09	5	Monday	420	2021-06-20	80	2021-07-31	293
2	2021-06-10	5	Saturday	464	2021-06-27	85	2021-08-31	243
3	2021-06-12	17	Sunday	515	2021-07-04	97	2021-09-30	258
4	2021-06-13	27	Thursday	397	2021-07-11	62	2021-10-31	316
5	2021-06-14	21	Tuesday	509	2021-07-18	76	2021-11-30	261
6	2021-06-15	6	Wednesday	445	2021-07-25	77	2021-12-31	171

Fig. 3 Data Classification

분석 결과, 예시로 활용된 카페의 경우 일요일(515건)과 화요일(509)이 평균 대비 혼잡한 것으로 나타났다. 월간 데이터의 경우 휴가철인 7월부터 점차 늘어나 11월까지 지속적으로 증가하였다. 주간 데이터의 경우 최근의 증감추세를 파악할 수 있도록 최근 4주간의 데이터를 활용하여 변화를 확인할 수 있도록 하였으며 7월 말까지 점차 증가하는 추세를 보였다. 각각의 날짜별 기준을 평균하고 차이를 활용하여 혼잡도 정보를 Fig. 4와 같이 설정할 수 있다.

Posting trends in the last 4 weeks			
-4	-3	-2	-1
-34%	19%	3%	12%

Ratio of Posting compared to week average						
Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
-6%	14%	-1%	-11%	-13%	4%	15%

Ratio of Posting compared to monthly average					
1	2	3	4	5	6
-27%	-46%	-19%	-11%	26%	-10%
7	8	9	10	11	12
-8%	16%	23%	51%	25%	-18%

Fig. 4 Result of Posting Data Analysis

이상에서 살펴본 바와 같이 게시글의 데이터를 분석하여 요일별, 주간별, 월간별 데이터를 통해 최종 혼잡도를 예측하는데 사용한다.

3.5 이미지를 통한 혼잡도 분석

본 연구에서는 이미지 내에 얼마나 많은 사람이 있는지 파악하기 위해 이미지 분석을 실시한다. 이미지 내 사람이 많다면 관광지에 사람이 많이 방문한 것으로 파악되어 혼잡도를 예측하는데 활용할 수 있다.

이미지 분석은 파이썬과 컴퓨터 비전 분야의 대표적인 오픈소스 라이브러리인 OpenCV를 활용하였다. OpenCV에서 제공되는 여러 기능 중 기울기 벡터를 기준으로 주로 전신의 사람을 추출하는데 사용되는 HOG(Histogram of Oriented Gradient)를 통해 이미지 내 사람을 파악하였다.

작성한 코드가 Fig. 5와 같이 전신의 사람을 앞모습과 뒷모습 모두 파악하고, 피사체 뒤편에 있는 사람도 정상적으로 인식하는 것을 확인할 수 있다.



Fig. 5 Whole Body Extraction

이미지 분석 단계에서도 게시글 분석 단계와 동일하게 각각의 이미지 내 포함된 사람의 수를 요일별, 주간별, 월별로 분리하여 혼잡도 분석을 실시하였다.

3.6 이미지를 통한 방문자 특성 분석

다음으로 관광지를 방문하는 사람들의 특성을 파악하기 위해 이미지를 분석하였다. 파이썬과 OpenCV를 활용하여 방문자 특성 중 성별과 연령대를 추출하였고, Haar Cascade 기반의 Front Face 탐지 모델이 사용되었다. 방문자는 성별과 연령대를 기준으로 8개 집단(유아, 영아, 어린이, 10대~20대 초반, 20대 중반~30대 초반, 30대 중반~40대 초반, 40대 중반~50대, 60대 이상)으로 구분하였다.

작성한 코드가 이미지 내 사람의 성별과 연령대를 적절하게 구분해내는 것을 Fig. 6에서와 같이 확인할 수 있었으며 3,138개의 포스팅에서 2,261명의 성별과



Fig. 6 Example of Characteristics Extraction

연령대 정보를 수집할 수 있었다.

성별과 연령대는 사람의 전면 이미지가 필요하므로 사람 수를 파악하는 이미지 분석에 비해 인식된 사용자가 적었으며, Confidence Level이 0.7 이상인 경우에만 실험 결과에 반영하였다. 분석 결과 영아 136명, 유아 100명, 어린이 143명, 10대~20대 초반 159명, 20대 중반~30대 초반 422명, 30대 중반~40대 초반 218명, 40대 중반~50대 48명, 60대 이상이 45명으로 나타났다. 이를 통해 연구에 사용된 카페의 경우 여성이 약 62%, 20대에서 30대 초반이 33%, 30대 중반에서 40대 초반이 17%로 20대에서 40대 초반의 여성이 주요 방문객임을 알 수 있었다.

3.7 최종 혼잡도 산출

앞서 도출된 데이터를 활용하여 이용자에게 관광지 의 최종 혼잡도를 제시한다. 혼잡도 점수 C 는 아래 수식과 같이 계산된다.

$$C_P = w_{P,d} \cdot C_{P,d} + w_{P,m} \cdot C_{P,m} \quad (1)$$

$$C_I = w_{I,d} \cdot C_{I,d} + w_{I,m} \cdot C_{I,m} \quad (2)$$

$$C = \alpha \cdot C_P + \beta \cdot C_I \quad (3)$$

where,

$$0 \leq w_{P,d}, w_{P,m}, w_{I,d}, w_{I,m}, \alpha, \beta \leq 1$$

$$w_{P,d} + w_{P,m} = 1, w_{I,d} + w_{I,m} = 1, \alpha + \beta = 1$$

수식에서 w 는 가중치를 의미하며, 포스팅 데이터 기반의 점수는 P 로, 이미지 기반의 점수는 I 로 표기한다. 수식 (1)에서 $C_{P,d}$ 는 혼잡도를 파악하고자 하는 일자에 해당되는 요일에 업로드된 포스팅 수가 분석

기간 내 업로드된 전체 포스팅 수에서 차지하는 상대적인 비율을 의미한다. $C_{P,m}$ 은 혼잡도를 파악하고자 하는 일자가 속하는 월의 포스팅 수가 분석 기간 내 전체 포스팅 수에서 차지하는 상대적인 비율을 뜻한다. 수식 (2)는 업로드된 이미지에 포함된 사람의 수를 기준으로 혼잡도를 산출한다. $C_{I,d}$ 는 혼잡도를 파악하고자 하는 일자에 해당되는 요일에 업로드된 포스팅 내 이미지에 포함된 사람 수의 합이 분석 기간 내 업로드된 전체 포스팅 이미지에서 추출된 사람 수의 합에서 차지하는 상대적인 비율을 뜻하며, $C_{I,m}$ 은 혼잡도를 파악하고자 하는 일자가 속하는 월에 업로드된 포스팅 내 이미지에 포함된 총 사람 수가 분석 기간 내 업로드된 전체 포스팅 이미지에서 추출된 총 사람 수에서 차지하는 상대적인 비율을 의미한다. 최종 혼잡도 점수 C 는 수식 (1)과 (2)의 가중합으로 산출된다. 수식 (3)에서 α 는 포스팅 데이터 측면의 가중치, β 는 이미지 측면의 가중치를 의미한다.

사용자가 특정 일자에 대한 최종 혼잡도를 확인하고자 할 경우, 시스템은 사용자가 요청한 일자의 요일과 월을 확인하게 된다. 이후 요일과 월별 혼잡도 데이터를 추출하고 정해진 가중치를 기반으로 포스팅 데이터 기반의 혼잡도와 이미지 기반의 혼잡도를 앞서 제시한 것처럼 추출한다.

예를 들어 이용자가 8월 10일 카페의 혼잡도를 예상하고 싶다면 시스템은 혼잡도 예측을 위해 포스팅 데이터와 이미지를 추출하여 분석한다. 8월 10일은 수요일이므로 $C_{P,d}$ 값을 사용하며, 예시에서 산출된 수요일의 혼잡도는 0.45로 나타났다. 8월의 혼잡도 수치는 $C_{P,m}$ 에 해당하며 본 연구에서는 0.68(C_P)로 산출되었다. 이후 요일과 월에 대한 반영비율인 가중치 $w_{P,d}$, $w_{P,m}$ 를 활용하여 포스팅 기반 혼잡도를 예측한다. 본 연구에서는 초기값으로 각각 0.5의 가중치를 사용하였고 포스팅 데이터 기반의 혼잡도는 0.58로 산출되었다. 같은 방법으로 이미지 기반의 혼잡도를 파악한다. 이미지에서 추출된 사용자 수를 통해 수요일의 이미지 기반 혼잡도인 0.11($C_{I,d}$)과 8월의 이미지 기반 혼잡도인 0.53($C_{I,m}$)을 도출하고 가중치를 활용하여 이미지 기반 혼잡도인 0.32(C_I)를 산출한다. 이후 포스팅 데이터 기반의 혼잡도와 이미지 기반의 혼잡도를 가중 합산하여 최종 혼잡도를 산출한다. 본 연구에서는 포스팅 데이터 기반 혼잡도와 이미지 기반 혼잡도 가중치 α , β 를 각각 0.5로 설정하였다. 이

를 통해 8월 10일 카페의 최종 예상 혼잡도는 0.5로 산출되었으며 평균적인 혼잡도를 보여줄 것으로 나타났다. 오픈한 지 1년이 경과하지 않은 신규 관광지의 경우 월별 혼잡도 정보를 파악할 수 있는 누적 데이터가 충분하지 않아 발생하는 Cold Start 문제에 대한 해결책이 필요하다. 즉 수식 (1), (2)에서 월별 혼잡도 변수 $C_{P,m}$ 과 $C_{I,m}$ 을 사용할 수 없으므로 월 가중치 $w_{I,m}$ 과 $w_{P,m}$ 을 0으로 설정하고, 포스팅 데이터와 이미지 내 사람 수 데이터의 최근 4주간 변화 추이를 반영하여 최종 예상 혼잡도를 예측할 수 있다. Fig. 7은 최종적으로 제공되는 혼잡도 분석 결과 사용자 인터페이스(UI: user interface) 예시이다.



Fig. 7 UI for Congestion Results

4. 시스템 검증

4.1 혼잡도 검증

본 연구에서 제안한 시스템을 검증하기 위하여 타 서비스와의 비교 분석을 실시하였다. 구글과 네이버에서 일부 유명 관광지에 대한 시간대별 혼잡도를 제공하고 있으나 대다수의 관광지에 대한 혼잡도 정보를 확인할 수 있는 서비스는 존재하지 않으며 특히 일별로 제공하는 국내 서비스는 없다. 제안 서비스의 검증을 위해 네이버 DataLab에서 제공하는 검색 데이터 통계를 활용하여 제안 시스템에서 도출된 요일별 혼잡도와 월별 혼잡도 데이터와의 비교를 통하여 검증을 진행하였다. DataLab으로부터 2021년

8월에서 2022년 7월까지 1년간의 데이터를 추출하였으며, 사이트별 건수는 원앤온리 24,125건, 팔공산 케이블카 2,533건, 뽕로로 파크(아쿠아 빌리지) 8,995건이다. Timestamp data를 이용해 월별, 요일별로 분류하여 비교 검증에 사용하였다.

제주도 소재 카페 원앤온리의 경우 Fig. 8과 같이 네이버 검색 시스템과 비교한 결과를 얻을 수 있었다. 요일별 결과에서는 91%, 월간 데이터의 경우 81%의 일치율을 보였다. 전체적인 그래프 또한 네이버의 데이터와 유사한 형태를 보이고 있다.

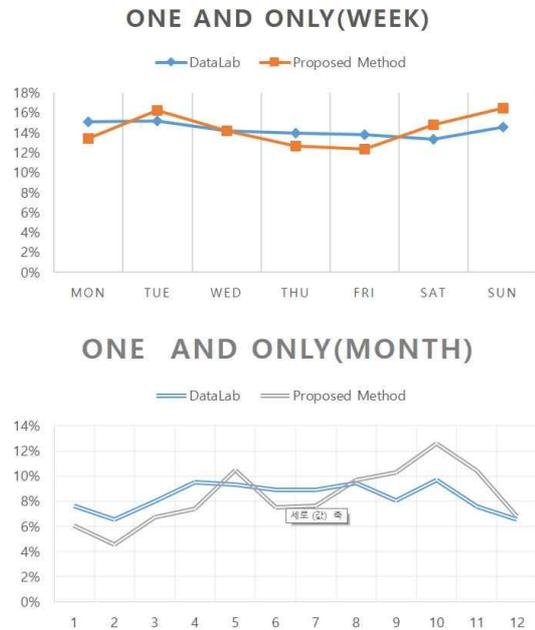


Fig. 8 Congestion Verification (Cafe)

다음으로 팔공산 케이블카를 추가로 분석하였다. 2021년 8월부터 2022년 7월까지의 데이터를 추출하였으며 총 1,400개의 포스팅과 이미지 분석을 통해 746명의 사람을 추출하였다. 팔공산 케이블카의 경우 Fig. 9와 같이 단풍이 드는 가을철부터 사용자가 급증하는 계절별 특성을 파악할 수 있는 관광지라 할 수 있다. 분석 결과 본 연구에서 예측한 혼잡도가 계절적인 특성을 잘 반영하고 있음을 확인할 수 있었다.

마지막으로 최근 하루 포스팅 수가 수십 개가 넘을 정도로 인기가 많은 경주의 뽕로로 아쿠아 빌리지를 대상으로 분석하였다. 수집된 포스팅은 총

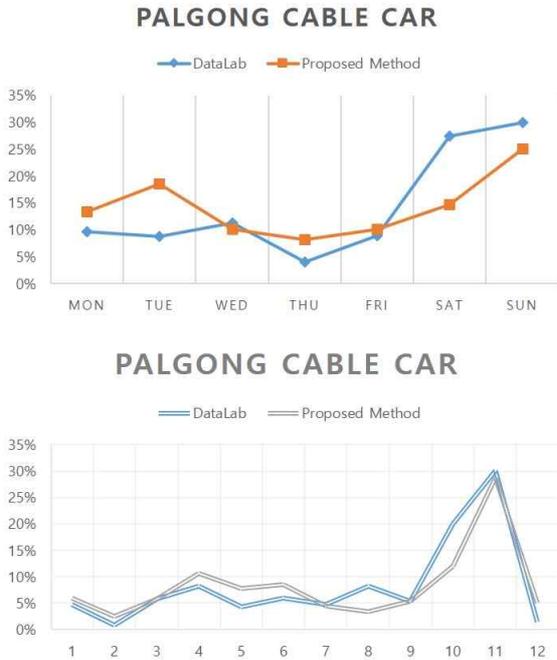


Fig. 9 Congestion Verification (Palgong Mt.)

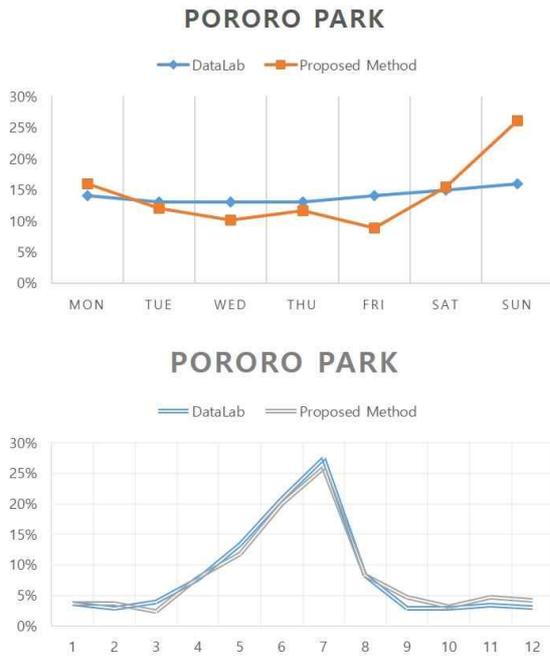


Fig. 10 Congestion Verification (Pororo Park)

4,224개이며 이미지 분석을 통해 추출된 사람의 수는 3,398명이다.

Fig. 10과 같이 월별 데이터는 거의 유사한 형태로 나타난 것을 확인할 수 있었다. 반면 요일별 데이터는 다소 차이를 보였는데 이는 해당 관광지의 예약 및 정보 확인 등 작업을 위해 검색한 사용자들로 인한 것으로 보인다. 이러한 결과는 검색어 기반의 혼잡도 시스템이 갖는 한계로서 실제로 방문하지 않았으나 검색 횟수가 증가하였기 때문이다. 반면 본 연구에서 제안한 방법으로 추출된 데이터의 경우 실제로 방문한 사용자가 업로드한 포스트를 대상으로 분석하였기 때문에 더욱 정확한 데이터라 할 수 있다. 방문자의 분포를 볼 때 제안 방법으로 추출한 데이터가 더욱 정확한 것으로 보인다.

4.2 방문자 특성 검증

이미지 분석을 통해 추출한 방문자 특성 검증을 위하여 DataLab에서 제공하는 데이터를 활용하였다.

DataLab에서는 일정 수준 이상의 검색어가 있고 인기가 있는 관광지에 한하여 사용자의 연령대와 성별을 제공한다. 본 연구에서 도출한 결과와 DataLab에서 제공한 데이터를 비교하여 검증하였다.

가장 먼저 연구에 사용된 원앤온리 카페의 분석 결과는 Fig. 11과 같다. DataLab에서는 10대, 20대, 30대와 같이 연령대를 기준으로 하고 있으며, 본 연구에서는 유아, 영아, 어린이, 10대~20대 초반, 20대 중반~30대 초반과 같은 형태로 구분하여 정확히

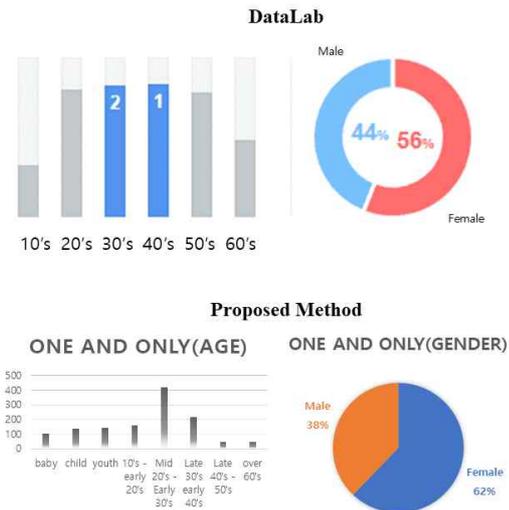


Fig. 11 Visitor Characteristics (Cafe)

일치하는 분류는 아니지만 전체적인 인구의 분포는 유사한 것으로 나타났으며, 성별 특성도 유사하게 나타난 것을 확인할 수 있다.

다음으로 팔공산 케이블카의 방문자 특성을 비교해보았다. Fig. 12에 나타난 바와 같이 30~40대의 방문율이 가장 높다는 것을 확인할 수 있었다. 다만 50대부터 SNS의 사용자가 매우 적기 때문에 50세 이상의 방문객에 대한 예측은 상대적으로 불충분한 것으로 보인다. 이는 SNS를 활용한 본 연구의 특성으로 SNS 사용자가 적은 연령대의 활동을 제대로 파악하지 못한다는 한계가 있다. 이와는 대조적으로 검색 시스템의 경우 10세 이하 연령대에 대한 검색어 통계가 매우 적기 때문에 예측이 불가하지만 본 연구에서는 이미지 분석을 통해 10세 이하 어린이의 특성까지 파악할 수 있다는 장점이 있다.

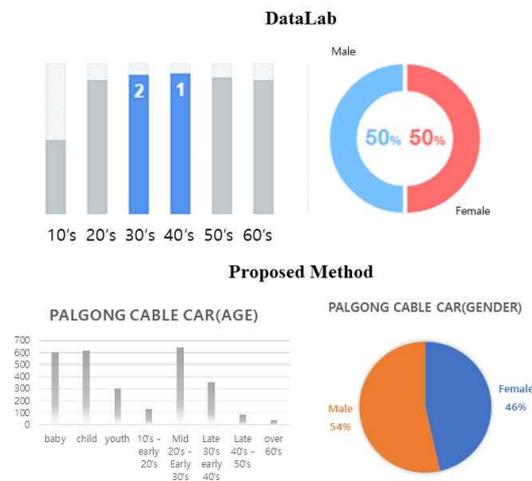


Fig. 12 Visitor Characteristics (Palgong Mt.)

이를 검증하기 위하여 10세 이하의 방문객이 많은 뽀로로 아쿠아 빌리지에 대한 사용자 특성을 분석한 결과는 Fig. 13과 같다.

분석 결과 DataLab에서는 추출되지 않은 영·유아 방문자가 매우 많은 것으로 나타났다. 이를 통해 검색어 기반의 시스템보다 낮은 연령대에 대해서도 보다 정확한 데이터를 수집할 수 있다는 것을 확인할 수 있었다.

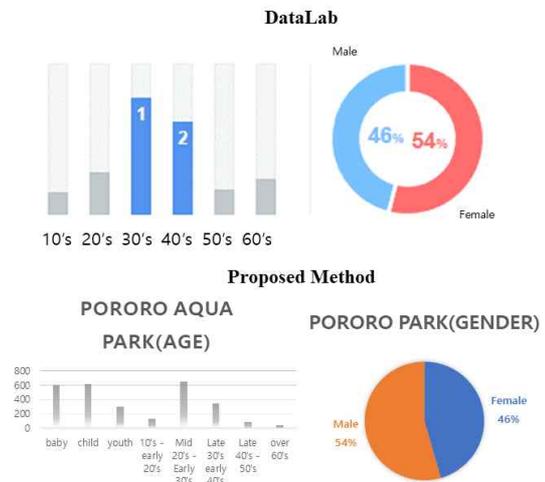


Fig. 13 Visitor Characteristics (Pororo Park)

5. 결론

본 논문은 SNS상에 있는 데이터를 활용하여 다양한 관광지에 대한 혼잡도를 예측하고 사용자에게 제공하는 시스템을 제안하였다. SNS에 업로드되는 포스팅과 이미지를 활용하여 사용자에게 혼잡도를 제공해주고자 하였으며, 포스팅 데이터 중 업로드 시간을 활용하여 분석을 실시하였다. 혼잡도는 업로드된 시간을 기준으로 월별, 요일별, 주간별 데이터를 사용하여 산출하였다. 이미지 분석의 경우 이미지 내 포함된 사람의 수를 파악하여 분석에 활용하였으며, 이미지 내 사람의 수는 OpenCV를 통해 추출하였다. 추출된 사람의 수를 월별, 요일별, 주간별 데이터로 분류하고 혼잡도를 산출하였다. 이후 포스팅 데이터의 혼잡도와 이미지 기반 혼잡도 데이터를 기반으로 최종 혼잡도를 산출하는 방식을 제안하였다. 방문자 특성을 파악하기 위하여 이미지 내 사람의 연령과 성별을 분석하였고 OpenCV의 Haar Cascade 기반 Front Face 탐지 모델을 사용하였다. 이를 통해 관광지 내 방문객의 연령대와 성별을 확인할 수 있었다.

제시된 시스템의 검증을 위하여 네이버 DataLab 시스템의 검색량 통계를 활용하였으며, 제주도의 인기 카페인 원앤온리, 팔공산 케이블카, 경주 뽀로로 아쿠아 빌리지 데이터를 사용하였다. 분석 결과 혼잡도의 경우 네이버 DataLab

의 검색어 통계와 매우 유사한 것을 확인할 수 있었다. 반면 사용자 특성 분석의 경우 SNS 데이터를 통한 분석이 50~60대 방문자 정보를 정상적으로 반영하지 못하는 한계가 있음을 알 수 있었다. 반대로 검색어 통계의 경우 10대 이하에서 정상적인 데이터의 추출이 불가능하지만 본 연구에서 제안한 시스템은 10대 이하에서도 사용자 특성을 추출할 수 있어 상대적으로 뛰어난 결과를 보여주었다.

본 연구는 SNS를 통해 관광지를 검색하고 정보를 획득하는 젊은 층이 흔히 핫플레이스라는 곳에 한꺼번에 몰리는 점을 개선하고자 혼잡도 정보를 제공하는 방법을 제시하였다.

현재 검색 포털에서도 혼잡도를 제공하고는 있지만, 그 대상이 내부적으로 설정한 조건에 부합하는 일부 대형 관광지로 한정된다. 반면 본 연구에서 제안한 방법을 사용하면 SNS상에 업로드된 데이터가 있는 모든 관광지에 대한 혼잡도 분석이 가능하며 규모가 작거나 오픈한 지 얼마 되지 않은 곳이라도 혼잡도 예측이 가능하다.

데이터 측면에서 본다면 검색 포털의 경우 혼잡도를 측정할 수 있는 모든 형태의 데이터를 제공하지 않고 극히 제한적인 일부 데이터만을 열람할 수 있으나 본 연구에서 제안하는 방식은 개인이 언제든지 수집할 수 있는 SNS의 데이터를 활용하여 혼잡도를 예측할 수 있다.

또한 검색 통계 기반 혼잡도의 경우 실제 방문하지 않는 사람이 단순 검색만 하더라도 혼잡도에 영향을 주게 되어 실질적인 상황을 반영하지 못하는 문제가 있지만, 본 연구에서는 사용자가 직접 위치 태그를 설정하거나 GPS에 기반한 게시물만을 입력으로 사용하여 실 방문자의 데이터를 토대로 산출된 혼잡도를 제공함으로써 보다 현실적인 정보를 제시한다는 장점이 있다.

향후 연구에서는 포스팅 내 포함된 태그와 텍스트 내용 등을 활용하여 더욱 개선된 시스템을 제시해야 할 것으로 보이며, SNS 데이터만으로는 분석이 어려운 50대 이상 방문자에 대해서도 수집이 가능한 포털 검색 통계와 함께 하이브리드 방식을 적용하여 혼잡도 예측의 정확도를 개선하기 위한 추가 연구가 필요하다.

References

- Han, Jinseok (2020). The Influence of Shop Crowdedness in Dining Franchise Liquor Shops upon the Perceived Values and Intent to Revisit, *Journal of the Korea Academia-Industrial Cooperation Society*, 21(5), 362-371.
- Hwang, YeonHee, Kim, YoungBerm and Lee, SangHo (2020). Effects of Real Name Using on SNS User's Behavior and Continuous Use Intention, *Journal of the Korea Industrial Information Systems Research*, 25(3), 83-97.
- Jeon, Su-Hyeon and Kwahk, Kee-Young (2016). Effects of Customers' Relationship Networks on Organizational Performance - Focusing on Facebook Fan Page -, *Journal of Intelligence and Information Systems*, 22(2), 57-79.
- Jun, Jae-Kyoon and Kim, Dong-Jin (2021). The Effects of Untact Tourism Motivation on Satisfaction and SNS WOM Intention, *Journal of Tourism and Leisure Research*, 33(12), 213-234.
- Kang, Youn-Young and Chon, Hae-Jung (2021). The Effect of Beauty SNS Exposure Information on Hair Style Preference and Purchase Intention, *Journal of the Korean Society for Wellness*, 16(2), 111-116.
- Kim, Dae-Seok, Bang, Won-Gi and Seo, Young-Wook (2022). A Study on Tourism SNS Benefits Effecting on Behavioral Intention(Reuse Intention, Satisfaction) through Tourism Flow : The Moderating Effects of SNS Type, *Journal of the Korea Entertainment Industry Association*, 16(3), 29-40.
- Kim, Dae-Seok and Choi, Eun-Gyo (2021). The Effect of Tourism SNS Service Quality on Perceived Usefulness, Tourism Flow and Tourism Experience, *Journal of Tourism and Leisure Research*, 33(7), 41-66.
- Kim, Dae-Seok and Lee, Eun-Ha (2021). A Study of Tourism Destination Storytelling on the

- Tourism Value and Relationship Immersion of Travel Agency: Moderating Effect of SNS Trust, *Journal of the Korea Contents Association*, 21(11), 503-517.
- Kim, Gyu-Sig and Oh, Day Young (2021). Comparison of Effectiveness between Traditional Media and SNS in Policy PR of Local Governments: Survey Result of Public Information Officials in 31 Counties of Gyeonggi-do, *Journal of The Korean Regional Development Association*, 33(2), 117-141.
- Kim, jihyo (2021). Instagram Feminism: SNS Reputation and Digital Feminism, *Journal of Korean Women's Studies*, 37(4), 119-154.
- Kim, Y. J. and Jang, A. (2021). A longitudinal study of sales promotion on social networking sites (SNS) in the lodging industry, *Journal of Hospitality and Tourism Management*, 48(June), 256-263. <https://doi.org/10.1016/j.jhtm.2021.06.016>.
- Kim, Young-Ha and Park, Hyun-Jee (2013). The Effect of Reuse Intention and Customer Satisfaction to Customer Value in Tourism Social Network Service, *Journal of Tourism and Leisure Research*, 25(5), 183-198.
- Kwak, Dae-young (2021). The Impact of SNS Tourism Information on Urban Image and Visit Intention: Focusing on Moderated Mediation Effects of Tourism Information Presence, *Korean Journal of Urban Studies*, (20), 113-146.
- Lee, Do-Yul and Kwahk, Kee-Young (2021). A Study on the Factors Affecting Photo Sharing Users on SNS: Focusing on Social Capital and Visual Sociology, *Korean Management Review*, 50(5), 1217-1254.
- Lee, Hyewon, Park, Hyeyoon and Park, Soyeon. (2022). Study on the Effect of Influencer on Advertising Attitudes and Brand Transformation Behavior: Focusing on Air Tourism related SNS Influencer. *Journal of Korea Culture Industry*, 22(1), 77-88.
- Lee, Jin Seo and Mo, Jeong Hee (2022). The Effects of SNS Advertisement Attributions on Behavior Intention in Selection of Beauty Salons, *Journal of the Korean Society Design Culture*, 28(2), 289-301.
- Lee, Jong-Hwa, Le, Hoanh Su and Lee, Hyun-Kyu (2016). Research on Methods for Processing Nonstandard Korean Words on Social Network Services, *Journal of the Korea Industrial Information Systems Research*, 21(3), 35-46.
- Lee, Jong-Hwa (2018). Building an SNS Crawling System Using Python, *Journal of the Korea Industrial Information Systems Research*, 23(5), 61-76.
- Lee, Yeon-Jung (2022). A Study on the Effect of SNS Information Characteristics on Value, Satisfaction and Reuse Intention of Restaurant Customers, *Culinary Science & Hospitality Research*, 28(4), 119-132.
- Li, J. G. T., Kim, J. O. and Lee, S. Y. (2009). An empirical examination of perceived retail crowding, emotions, and retail outcomes, *Service Industries Journal*, 29(5), 635-652. <https://doi.org/10.1080/02642060902720121>.
- Li, X., Pan, B., Law, R. and Huang, X. (2017). Forecasting tourism demand with composite search index, *Tourism Management*, 59, 57-66. <https://doi.org/10.1016/j.tourman.2016.07.005>.
- Machleit, K. A., Kellaris, J. J. and Eroglu, S. A. (1994). Human versus spatial dimensions of crowding perceptions in retail environments: A note on their measurement and effect on shopper satisfaction, *Marketing Letters*, 5(2), 183-194. <https://doi.org/10.1007/BF00994108>.
- Manning, R. E. and Ciali, C. P. (1980). Recreation Density and User Satisfaction: A Further Exploration of the Satisfaction Model, *Journal of Leisure Research*, 12(4), 413-420.
- Oh, Kyung-Heon (2022). Effects of SNS Marketing by Beauty Service Customers on Word of mouth effect and purchase behavioral intention,

Journal of the Korea Academia-Industrial Cooperation Society, 23(6), 252-261.

Quan, Wei and Han, Hee-Sup (2019). How crowdedness affects Chinese customer satisfaction at Korean restaurants?, *Korean Journal of Hospitality & Tourism*, 28(8), 161-177.

Saenz-de-Miera, O. and Rosselló, J. (2012). The responsibility of tourism in traffic congestion and hyper-congestion: A case study from Mallorca, Spain, *Tourism Management*, 33(2), 466 - 479. <https://doi.org/10.1016/j.tourman.2011.06.015>.

Son, Jung-Il and Heo, Chul-Moo. (2018). Effect of the search index on Purchase behavior by SNS channel attributes, *Journal of Digital Contents Society*, 19(8), 1535-1544.

Song, Seung-A and Shin, Hyung-Deok (2022). Presentation of Self and SNS Posting Styles: Focusing on Goffman's Impression Management Framework, *Journal of the Korea Contents Association*, 22(4), 284-291.

Xie, G., Qian, Y. and Wang, S. (2021). Forecasting Chinese cruise tourism demand with big data: An optimized machine learning approach, *Tourism Management*, 82(October 2019), 104208. <https://doi.org/10.1016/j.tourman.2020.104208>.

Yang, Sung-Soo, Oh, Eun-Ji and Jeong, Chul (2019). The Effect of Expectancy on Crowding, Crowding Perception, Satisfaction, and Behavioral Intention for Winter Tourist on Island : A Case of visitors to u-do Island, *Journal of Tourism Studies*, 31(3), 121-140.

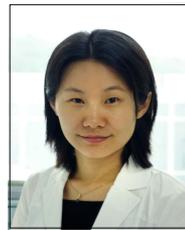
Yang, Y., Guo, J. and Sun, S. (2021). Tourism demand forecasting and tourists' search behavior: evidence from segmented Baidu search volume, *Data Science and Management*, 4(September), 1-9. <https://doi.org/10.1016/j.dsm.2021.10.002>.

Yang, Young Hwan and Kang, Min Jung (2022). A Study on the Effect of the SNS Characteristics on Public Organization's Image and Policy Participation, *Journal of Business Convergence*, 7(1), 55-64.



이 상 훈 (Sang Hoon Lee)

- 대구대학교 정보공학과 공학사
- 대구대학교 경영학과 경영학사
- 대구대학교 컴퓨터정보공학과 공학석사
- 대구대학교 정보공학과 공학박사
- (현재) 대구대학교 정보통신대학 컴퓨터정보공학부 강사
- 관심분야: 추천시스템, 지식재산권, Web 3.0



김 수 연 (Su-Yeon Kim)

- 종신회원
- 포항공과대학교 수학과 이학사
- 숭실대학교 정보산업학과 이학석사
- 포항공과대학교 산업경영공학과 공학박사
- (현재) 대구대학교 정보통신대학 컴퓨터정보공학부 교수
- 관심분야: 지능형 시스템, 기술경영, 지식재산권