

## Original Article

<https://doi.org/10.12985/ksaa.2020.28.2.036>  
ISSN 1225-9705(print) ISSN 2466-1791(online)

## 빅데이터 분석을 활용한 공항 혼잡도 분석

## - 김포공항 사례를 중심으로 -

김진아\*, 김진기\*\*

## Airport Congestion Analysis with Big Data Analysis

## - The Case of Gimpo Airport -

Jin Ah Kim\*, Jin Ki Kim\*\*

## ABSTRACT

This study is designed to help customers use more comfortable airports by predicting congestion and congestion times by identifying the traffic routes of passengers in the airport building by day of the week and time by using Wi-Fi sensor collectors, one of the IoT technologies. Analysis of passenger traffic analysis data showed that the most congested time zones were from noon. to 2p.m. for all facilities, which could be used to improve major facilities. Regression analysis of factors affecting congestion found that self-check-in reduces congestion and check-in counters increases congestion. These findings will provide important implications for operations, including congestion management at airports.

**Key Words** : Airport Congestion (공항 혼잡), Traffic Flow(여객 흐름), Congestion(혼잡도), Staying Time(체류 시간), Big Data(빅데이터)

## 1. 서 론

최근 몇 년 전부터 우리는 각종 미디어 매체에서 “제4차 산업혁명 시대에 접어들었다”라는 말을 자주 접하고 있다. 현재 우리 사회에서 제4차 산업혁명이 진행되고 있는지, 그리고 그렇다면 기존과는 어떤 차이가 있는지, 그리고 그 핵심기술은 무엇인지, 이를 통해 우리 사회에 영향을 줄 수 있는 것은 무엇인지에 대한 답변이 필요할 것이다.

『제4차 산업혁명』이라는 용어는 2016년 다보스 세

계경제포럼에서 클라우드 슈바프 의장이 언급하면서 사용되기 시작되었고, 포럼에서 논의되었던 여러 사례를 정리하여 발표한 “The Fourth Industrial Revolution”을 발간하면서 전 세계로 내용이 확산되었다(Schwab, 2017). 이러한 현상은 최근 경제환경에서 새롭게 등장하고 있는 여러 현상을 설명하는 데 도움이 되었기 때문이다. 제4차 산업혁명은 지금까지 진행되었던 과거의 산업혁명과는 질적으로 큰 차이가 있기 때문이다. 기존의 산업혁명들은 인간이 자신의 육체노동을 대신할 수 있는 새로운 기술, 예를 들면 제1차 산업혁명은 18세기 증기 기반에 의한 기계화 혁명, 제2차 산업혁명은 전기에너지 기반에 의한 대량생산 혁명, 제3차 산업혁명은 컴퓨터와 인터넷 기반의 지식정보 혁명이었으나, 제4차 산업혁명은 인간이 가지고 있는 고유 영역인 지능적인 일들을 기존에 가지고 있는 기술들을 고도화하여 융·복합하는 기술에 기반을 두기 때문이

Received: 22. Mar. 2020, Revised: 21. Jun. 2020,

Accepted: 21. Jun. 2020

\* 한국공항공사 항로시설본부 인천항공교통시설단 시스템정보부 부장

\*\* 한국항공대학교 경영학부 교수

연락처 E-mail : kimjk@kau.ac.kr

연락처 주소 : 경기도 고양시 덕양구 항공대학로 76

다. 제4차 산업혁명에서 가장 중요한 핵심 키워드는 “융·복합”이다.

제4차 산업혁명을 이끄는 기술들로는 AI, 클라우드, IoT 센서, 빅데이터 등이 있다. 본 연구는 이 중에서 핵심기술이 빅데이터 기술을 더 자세히 살펴보고자 한다. 왜냐하면, 제4차 산업혁명을 가속할 수 있었던 그 중심축이 바로 빅데이터를 이용한 융·복합 기술이기 때문이다. 정보 기술의 발전 속도가 빨라지면서 더불어 데이터의 양이 기하급수적으로 많아짐에 따라 이 데이터의 양에 “빅”이라는 단어를 붙일 수 있을 만큼의 상태가 되면서 데이터의 가치를 다시금 생각하게 되었기 때문이다. 이러한 데이터를 가치 있게 만들기 위해 기계학습, 인공지능, 딥러닝 등의 기술을 사용하게 되었다.

정부에서도 적절한 정책 수립과 대응을 위해 대량의 데이터 분석·활용을 통해 새로운 가치를 창출하고, 혁신을 주도하는 빅데이터 플랫폼을 미래성장 동력으로 선정하였다.

그렇다면 많은 여객이 이용하는 공항에서는 이러한 빅데이터를 어떻게 활용할 수 있을까? 현재 김포, 김해, 제주 등 주요 공항은 날이 갈수록 공항 혼잡에 따른 지연 문제가 연일 화두가 되고 있다. 공항 혼잡에 따른 지연은 고객의 불만뿐만 아니라, 경제적으로도 큰 손실이 야기된다. 본 연구에서 공항 혼잡은 에어사이드 부분, 즉 항로의 혼잡, 주기장 수용량의 한계에 따른 혼잡 등도 있지만 청사 내 혼잡도 포함이 된다. 이런 혼잡도를 해결하는 방법 가운데 하나는 공항 시설과 같은 인프라를 확장하는 것이다(박정근, 2010; 배열, 2018).

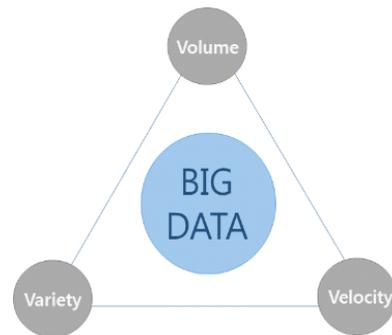
하지만 그 방법은 시간, 비용 및 주변 여건상 쉽게 해결할 수 없는 부분이다. 이에 공항 시설을 효율적으로 활용하여 혼잡도를 완화하여 공항을 찾는 고객이 심리적으로 느끼는 혼잡도를 축소하는 방안에 대한 필요성이 제기되었다.

단기적인 혼잡도 감소방안으로는 (1) 공항 시설을 재배치하거나 처리 프로세스 개선, (2) 탑승객의 흐름 동선을 최소화하고 효율화 (3) 탑승객이 공항에 도착해서 출발할 때까지의 단계별 처리 시간을 간소화가 제안되었다(강윤주, 2017).

본 연구는 공항 청사 내 여객의 흐름을 파악하여 공항의 효율성을 높이는 방안을 강구하고, 이에 따르는 문제점을 파악하고자 한다.

## II. 이론적 배경

빅데이터란 정보화 사회에서 생성되고 가공되는 수



출처: Turban, et al. (2018)

Fig. 1. 빅데이터의 특징

많은 데이터로, 그 규모가 매우 크고, 생성 주기가 짧으며, 그 형태도 정형화된 수치 데이터뿐만 아니라, 문자나 영상 등 반정형, 비정형 데이터를 포함하는 대규모의 데이터를 말한다. 빅데이터 환경은 과거보다 데이터의 양이 급증했다는 그것뿐만 아니라, 데이터의 형태도 매우 다양해져서 사람들의 행동은 물론 다양한 소셜 정보(SNS 등)를 통한 개인의 성향, 생각까지 분석하고 예측할 수 있다(서현진, 2017; 안효선, 2017).

### 2.1 빅데이터 분석

디지털 경제가 빠르게 확산함에 따라 우리 사회는 규모를 가늠할 수 없을 만큼의 많은 정보와 데이터가 생산되는 ‘빅데이터(Big Data)’ 환경에 직면하게 되었다. 정보화 시대에서는 과거 아날로그환경과 비교하면 정보통신기술과 기기의 발전으로 디지털 대응량 정보임을 강조하는 “빅”이라는 용어를 붙여 “빅데이터”라고 정의하였다. 이는 일반적인 정보처리 기술로 처리할 수 없는 크기로 현재 수십 테라바이트에서 수 페타바이트 이상 그 양이 매우 방대하다.

요즘 대다수의 TV 광고를 보면 IoT와 빅데이터, VR 기술을 활용한 광고들이다. 개인의 구매 취향, 구매 이력 등의 데이터를 활용하여 고객이 주문하기 전에 알아서 주문 여부를 알려준다거나, 고객의 예상 일정, 날씨별 고객의 취향에 맞는 음악 등을 틀어주기도 한다.

쇼핑 부문에서도 변화가 나타나고 있는데, 데이터의 관점에서 본다면 과거에는 상점에서 물건을 살 때만 고객과 관련된 데이터가 기록되었다. 그러나 인터넷 쇼핑몰의 경우에는 구매자가 실제 구매 이력이 없을지라도 방문자가 돌아다니던 흔적이 자동으로 데이터로 저장된다. 어떤 상품에 관심이 있는지, 얼마 동안 어느

Table 1. Top 5 global companies in terms of market price

| 순위 | 2019년   |          | 2018년   |          | 2017년   |          |
|----|---------|----------|---------|----------|---------|----------|
|    | 업체명     | 시가 (억\$) | 업체명     | 시가 (억\$) | 업체명     | 시가 (억\$) |
| 1  | 마이크로소프트 | 8,226    | 애플      | 8,144    | 애플      | 8,000    |
| 2  | 애플      | 8,213    | 알파벳     | 7,749    | 알파벳     | 6,839    |
| 3  | 아마존     | 8,148    | 마이크로소프트 | 7,067    | 마이크로소프트 | 5,382    |
| 4  | 알파벳     | 7,990    | 아마존     | 6,922    | 아마존     | 4,791    |
| 5  | 페이스북    | 4,885    | 페이스북    | 5,574    | 페이스북    | 4,462    |

사이트에 머물렀는지 알 수 있다. 또한, 개인이 사용하는 IoT를 통해 고객의 취향, 사용주기, 생활 방식 등을 융·복합해서 분석하게 된다. 즉, 사람과 기계, 기계와 기계 등 서로 정보를 주고받는 IoT 기기의 사용 증가도 디지털 정보가 급격히 증가하게 되는 이유다.

세계적 기업들은 빅데이터를 적극적으로 활용하고 있다. 특히 검색 엔진과 전자상거래 기업은 방대한 고객 데이터를 분석해서 다양한 방식의 마케팅 활동을 하고 있다. 구글의 자동번역 시스템, 아마존 고 등의 추천시스템은 대표적이다. 공공 부문도 위험관리시스템, 서울시 올빼미 심야버스, 부동산 자료 공개 등 빅데이터를 적극적으로 활용하고 있다. 기업들은 데이터의 가치를 잘 인지하고 있어서 데이터 관련 분야에 적극적이다. 실제로 세계적 기업 시가 총액 Top 5에 있는 업체들을 보면 모두 데이터를 사용하는 기업임을 알 수 있다.

## 2.2 빅데이터 활용 사례

해외에서 빅데이터의 대표적 활용사례의 선두주자는 구글을 들 수 있다. 구글은 데이터의 정확도를 떠나서 일단 데이터가 쌓이면 쌓일수록 정보의 품질도 좋아진다는 것을 인터넷 검색 결과에서 보여주고 있는 기업이다.

구글은 많은 양의 데이터를 처리하기 위하여 분석 도구의 핵심인 맵리듀스와 구글분산파일시스템이라는 새로운 처리 기술을 개발하였고, 이 기술이 현재 빅데이터 분석이 발전하는 큰 원동력이 되었다(김재생, 2014).

아마존은 온라인 쇼핑몰의 선구자로서 고객 유형별 소비패턴을 분석하여 예상되는 구매 목록을 고객에게 권하는 서비스를 최초로 시도하였다. 1단계로 도서 구매 데이터를 분석함으로써 고객이 구매할 것이라고 예

상되는 도서를 추천하고, 이에 해당하는 할인 쿠폰 등을 지급하는 방식으로 빅데이터를 활용하였다. 현재는 도서를 뛰어넘어 다양한 쇼핑 품목으로 그 활용범위를 넓힌 상태이다.

요즘 세계적으로 관심을 받는 넷플릭스는 이용자의 영화 대여 목록을 기초로 하여 새로운 영화를 추천해주는 시네매치(Cinematch) 시스템을 개발하였다. 이 시스템이 바로 넷플릭스가 많은 이용자를 확보하는데 큰 견인차 구실을 하였다.

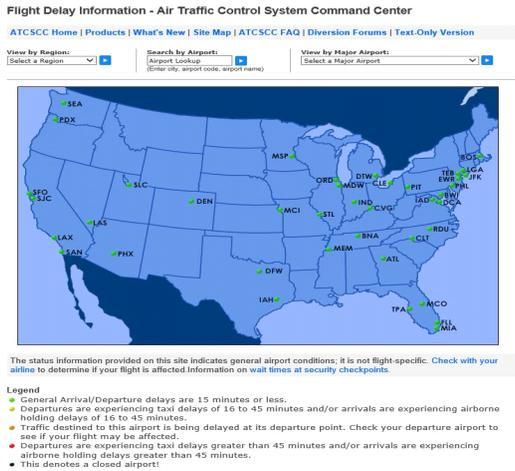
국내에서 빅데이터를 활용하는 사례를 살펴보면, 공항과 관계가 깊은 관광 분야에서는 외국인이 제주도를 관광할 때 다양한 정보를 주기 위한 생활 및 관광 밀착형 O2O 비컨 서비스 “땡땡 제주”가 있다. 이 서비스는 제주에서 여행객이 이동하는 동선에 따라 근거리 사물인터넷 기술인 비컨을 활용하여 근처 관광지 정보 및 맛집 정보 등을 알려주는 서비스이다(박로은, 2017).

이 외에도 공항 면세점이나 상업시설에 활용이 될 만한 것으로는 강남역 스포츠 매장에 Wi-Fi 신호를 분석하여 매장 운영을 분석한다거나, 수입품 판매점 매장을 철수하기로 하는 데에도 Wi-Fi 센서를 이용한 빅데이터를 활용하였다(김장수, 2017; 이영준, 2013).

또한, 공항에 매우 중요한 보안 분야에 사용할 수 있는 사례는 현재 은행 내부에서 Wi-Fi 신호를 스캔하여 분석함으로써 은행 영업장 행동 분석, 이상 감지 및 야간시간대의 침입 감지 등에 활용하고 있다(문대홍, 2016).

해외 공공 부문에서는 주로 국민의 안전과 편의에 주안점을 두고 있다. 예를 들면 싱가포르 정부에서는 재난 방재와 더불어 테러 감지, 전염병 확산과 같은 불확실한 미래를 대비하기 위해 국가위험관리시스템(Risk Assessment & Horizon Scanning: RAHS)을 2004년부터 지속해서 추진하고 있다. 또한, 미국 연방수사국(FBI)은 빅데이터를 활용한 DNA 색인 시스템을 구축하였다. 이를 통해 단시간에 범인을 검거하는 데 사용하고 있다. 이 외에도 미국 연방항공청(FAA)은 미국 내의 민간항공안전을 위한 중요한 임무를 맡은 정부 기관으로 항공 운항 관련 빅데이터를 활용하여 실시간으로 미국 전역에 운항 중인 항공기의 지역별 항공사별 정시 현황을 시각화하여 제공하고 있다(최문기, 2018).

반면, 국내 공공 부문의 빅데이터 활용은 이제 시작 단계에 불과하다. 빅데이터의 중요성과 필요성에 대해서는 이미 공감대가 충분히 형성되어 있으나, 우리나라



출처: Federal Aviation Administration (https://www.fly.faa.gov/flyfaa/usmap.jsp)

Fig. 2. Flight delay information-air traffic control system command center

의 특수상황으로 인하여 보안과 개인정보보호의 중요성이 강조됨에 따라 빅데이터의 발전은 아래래도 해외 성장 속도를 따라가기는 힘들다(김진수, 2016, 이종진, 2017).

국내 공공 부문 중 빅데이터 분석의 대표적인 성공 사례는 서울특별시에서 구축한 오페미 심야버스 노선 신설이다. 주로 강남지역의 카드 사용 및 교통수단 이용 현황을 분석하여 사람들이 주로 이동하는 패턴 분석에 의거 심야버스 노선을 신설 확대한 사례이다(김해, 2017; 한여희, 2018).

이 외에도 교통안전공단의 차종별 위험 운전 행동기준을 분석하거나 한국토지주택공사에서 빅데이터 수집을 통한 주택 산업을 전망하는 데 활용하고 있다(민이슬, 2019; 윤종현, 2017; 이상진, 2018).

### III. 공항 혼잡도

본 연구에서는 빅데이터의 특징 및 국내외 빅데이터 활용사례들을 참고로 하여 국내 공항에서 활용 가능한 방안을 제안하기 위하여 단기간에 많은 자료를 수집할 수 있는 공항의 Wi-Fi 센서 자료를 활용하였다. 이와 함께 Wi-Fi 센서 자료 분석 기준 및 분석 방안을 검토하였다.

공항 내 여객의 이동 흐름을 분석하여 공항 내부가 혼잡한지 여유로운지를 판단하기 위하여 공항 시설별 적절한 규모를 산정하기 위한 기준이 필요하다. 시설의

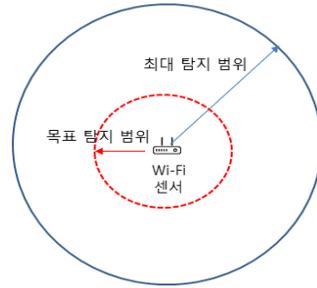


Fig. 3. Wi-Fi sensor detection range

적절한 규모를 산정하기 위해 국제적으로 통용되는 일반 기준에 따라 김포공항 시설에 적합한 기준을 마련하여 혼잡도를 측정하는 것이 필요하였다.

이를 위해 우선 혼잡도를 측정하기 위해 김포공항에 설치한 Wi-Fi 센서의 반경과 면적을 이용해 기준값을 아래와 같이 설정하였다.

#### 3.1 Wi-Fi 반경 설정과 면적 비율 산정

청사 내 시설별 적절한 Wi-Fi 스캐너의 센싱 면적을 도출하기 위해 시설별 반경과 면적 비율을 도출하였다.

청사 내 시설별 Wi-Fi 스캐너 센싱 반경은 실측을 통해 산정하고, 시설별 특성에 따라 3m, 5m, 10m, 20m를 적용하였다. Wi-Fi 스캐너는 원의 형태로 센싱되기 때문에 스캐너별 면적은 원의 넓이( $\pi r^2$ )를 가지게 된다.

실측을 통해 시설별 총면적에서 여객이 들어갈 수 없으므로 제외해야 하는 영역(예를 들면 직원 사무공간, 창문 밖, 기둥과 같은 부착물 등)을 확인하여 적정 비율을 산정하였다. 그 값은 Table 2의 면적 비율 (ratio)과 같다.

#### 3.2 국내선 여객 1인당 소요 면적

여객 1인당 소요 면적은 제3차 공항개발 중장기 종합계획 수립 조사의 여객터미널별 서비스 수준 및 미연방항공청에서 산정한 수준-I값을 기준으로 하였으며, 국내선 시설의 여객 소요 면적과 실측을 통해 많은 면적을 차지하는 캐리어나 카트를 가지고 들어가는 시설과 아닌 시설 등을 확인하여 시설별로 그 값을 조정하였다(건설교통부, 1999; 국토교통부, 2005). 이 부분은 변화에 따라 지속적인 조정이 필요한 값들이다.

국내선의 경우, 여객들의 캐리어 부피는 국제선보다

Table 2. Wi-Fi sensing space ratio and required space per passenger by airport facility

| 공항 시설     | 면적 비율 (Ratio, %) | 여객 1인당 소요 면적 (m <sup>2</sup> ) |
|-----------|------------------|--------------------------------|
| 게이트       | 80               | 2.1                            |
| 도착장       | 80               | 2.1                            |
| 신분검색대     | 60               | 2.1                            |
| 탑승구       | 40               | 2.1                            |
| 환승 게이트    | 80               | 2.1                            |
| 셀프 체크인    | 40               | 2.1                            |
| 체크인카운터    | 50               | 1.14                           |
| 보안검색대     | 17               | 1.2                            |
| 의자 구역     | 80               | 1.14                           |
| 복도(홀)     | 80               | 2.1                            |
| 화장실       | 33               | 1.14                           |
| 수화물 수취소   | 50               | 2.1                            |
| 면세점       | 20               | 0.8                            |
| 면세품 인도장   | 50               | 0.8                            |
| 식당        | 60               | 0.8                            |
| 편의점       | 60               | 0.8                            |
| 카페        | 60               | 0.8                            |
| 어린이 놀이방   | 80               | 1.0                            |
| 은행        | 60               | 0.8                            |
| 안내데스크     | 50               | 0.8                            |
| 약국        | 50               | 0.8                            |
| 우체국       | 60               | 0.8                            |
| 체크인 대기지역  | 100              | 2.1                            |
| 테이블 구역    | 60               | 2.1                            |
| TAX FREE  | 50               | 0.8                            |
| 기타 상업시설   | 50               | 0.8                            |
| 편의시설      | 80               | 1                              |
| 렌터카 구역    | 50               | 0.8                            |
| 기도실       | 60               | 0.8                            |
| 수유실       | 50               | 1                              |
| 유실물 센터    | 50               | 0.8                            |
| 수화물 보관소   | 20               | 0.8                            |
| 환전소       | 50               | 0.8                            |
| 입국 검역장    | 50               | 0.8                            |
| 입국 보안검색대  | 17               | 1.2                            |
| 직원용 보안검색대 | 50               | 1.2                            |
| 기타 출입문    | 80               | 2.1                            |

크지 않은 경우가 대부분이며, 카트를 이용하는 여객 또한 국제선보다 많지 않으므로 전체적으로 국내선 여객 1인당 소요 면적은 국제선의 1인 소요 면적보다 적

게 잡는 것이 바람직하다.

시설별로 여객이 캐리어나 카트를 들고 들어가는 구역과 아닌 구역을 구별하였다. 예를 들면, 여객이 처음 들어오는 게이트 구역과 체크인 구역은 캐리어나 카트를 가지고 이동하지만, 체크인카운터에서 캐리어를 맡긴 후에 가는 시설의 경우는 여객 1인당 차지하는 면적이 줄어들게 되므로 그 값을 다르게 적용하여 적정 소요 면적을 산정하였다.

각 기준을 적용한 결과는 Table 2와 같다.

### 3.3 혼잡도 구간 계산

1단계) 시설별 실측한 반경(r)을 이용하여 면적( $\pi r^2$ )을 구한 후 Table 2의 면적 비율을 곱하여 시설별 적정 면적을 도출함.

2단계) Table 2의 국내선 여객 1인당 소요 면적으로 시설의 적정 면적을 나눠 적정 인원을 도출함.

3단계) 시설별 도출된 적정 인원의 70% 이하를 혼잡도 수준으로 '원활'로 정하며, 130% 이상을 '혼잡', 그 사이 구간인 70%~130%를 '적정'으로 정의함.

빅데이터를 정확히 분석하기 위해서는 분석 전 단계의 과정도 매우 중요하다. 어떤 방법으로 분석할지 방법론을 결정한 후 시작해야 한다. 본 연구에서의 방법론은 Fig. 4와 같다.

한편, 혼잡도를 분석하기 위한 빅데이터 분석 절차는 Fig. 5와 같다.



Fig. 4. Big data analysis methodology



Fig. 5. Big data analysis procedure

### IV. 연구모형과 방법론

본 연구에서는 공항 설비의 혼잡도와 관련하여 기존의 연구 결과와 공항 운영을 통해 체득한 경험을 바탕으로 가설을 제안하였다.

일반적으로 공항에서 체류 시간(stay time)은 방문자가 많을수록 증가할 것으로 추정할 수 있다. 따라서 본 연구에서는 공항에서의 체류 시간은 방문자의 증가에 따른 지연(delay)으로 영향을 받을 것이라는 가설을 설정하였다. 따라서 기본 가설로, 여객의 체류 시간을 시설별 방문자 수의 함수로 판단하였다.

가설 1) 공항의 방문자 수( $V$ )는 공항 시설에서 방문자의 체류 시간에 영향을 준다.

일반적으로 특정 시설의 방문자 수는 포아송분포를 보인다. 즉, 특정 시점에 방문자의 수가 증가하고, 증가한 방문자는 혼잡(congestion)에 의해 특정 업무나 서비스 제공에 지연(delay)이 발생하게 된다. 공항에서도 같은 현상이 발생할 것이라는 추정에 따라 방문자 수의 증가에 따른 영향은 즉각적으로 나타나는 것이 아니라, 일정한 시간 이후에 지체 현상으로 나타난다. Nepal et al. (2019)의 연구에서도 방문자의 지연 현상을 반영한 지연함수(lag function)를 통해 반영하였다. 본 연구에서도 이러한 논리적 근거를 기반으로 지체 시간에 시간 지연(time lag) 현상이 나타난다는 것을 고려하였다.

가설 2) 공항의 방문자 수(1시간 이내 체류한 방문자 수,  $V_{t-1}+V_t$ )는 공항 시설에서 방문자의 체류 시간에 영향을 준다.

본 연구의 핵심 연구가설로 시설별로 체류 시간에 차이가 있을 것이라고 가정하였다. 즉, 시설의 특징에 따라 시설에서의 체류 시간에는 차이가 있을 것이다. 따라서, 기본 가설의 확장형 가설로, 개별 시설별로 체류 시간에 미치는 영향이 다를 것이라는 가설을 설정하였다.

가설 3) 공항의 시설(셀프체크인, 체크인카운터)별로 체류 시간에 차이가 있다.

본 연구에서는 IoT 기술 중 Wi-Fi 스캐너 센서를

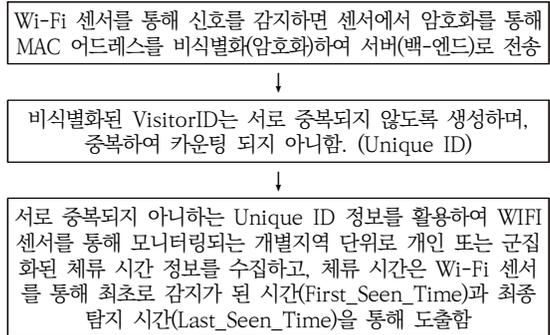


Fig. 6. Study design

활용하여 공항 방문자의 휴대전화기의 Wi-Fi 센서를 감지하여 여객 동선을 분석하기로 하였다. 또한, 여객 청사 내 혼잡도 해결 방안이 이 연구의 주요 과제이므로, 이를 위해 주요 시설별 혼잡 시간대와 체류 시간이 주요 변수로 선별하였다.

이 연구에서는 공항을 방문한 사람이 휴대전화기를 소지한 경우 휴대전화기의 Wi-Fi 센서를 스캔하는 방식으로 처리하였다. Wi-Fi 센싱에 의해 데이터를 분석하는 방법은 Fig. 6과 같다.

연구대상 자료는 2019년 6월 1일부터 6월 10일까지 총 10일간 김포공항 국내선 청사를 방문한 사람들이 휴대전화기를 소지한 경우 휴대전화기에서 발생하는 Wi-Fi 센서를 수집하여 이루어졌다. 이 결과를 빅데이터 분석 도구인 하둡과 ETL을 이용하여 분석하였다.

공항 내의 시설별로 Wi-Fi 센서에 접속된 단말의 수와 체류 시간을 측정하였고, 이를 분석 도구에 따라 분류하고, 데이터를 정제하였다. 이렇게 측정된 자료는 시설별 시간대별 방문자의 수와 체류 시간으로 측정되었다.

Wi-Fi에 접속한 데이터를 가지고 측정된 자료로, 데이터의 신뢰성에 문제가 있을 수 있다. 이를 제거하기 위해 추정가능한 특별사항에 대해서 이를 보정하였다. 그러나, 방문자 수에 대한 데이터는 이러한 보정작업을 통해 일정 정도 충분히 신뢰성을 확보하였다.

### V. 연구 결과

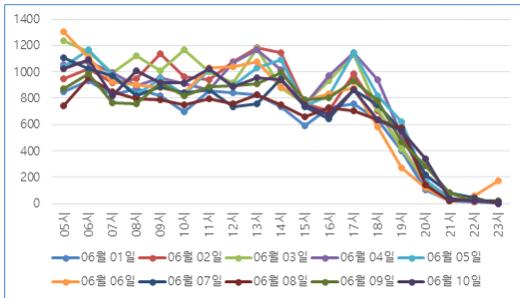
공항의 주요 시설별로 방문자 수와 체류 시간을 측정하였다. 셀프체크인의 경우, 방문자 수는 일자에 상관없이 고르게 분포되어 있다. 시간대는 일반적으로 첫 비행기를 타기 위한 5~6시대, 점심시간 대인 12~13시, 그리고 저녁 5시대가 방문객이 많다. 이 부분은 시

간대가 직원 교대 시간과 유사한 것을 고려하면 Wi-Fi 센서 데이터 수집할 때 상주 직원 자료가 포함되었을 가능성이 보인다.

6월 6일(현충일) 21시대가 소요 시간이 가장 많음을 알 수 있다. 셀프체크인은 전반적으로 21시대가 가장 소요 시간이 길다. 당일 김포공항 국내선이 A/C접속으로 58편 지연되고, 제주 강풍으로 11편이 결항하였다. 이 중 많은 편수가 오후 시간대라 영향을 미친 것으로 추측된다.

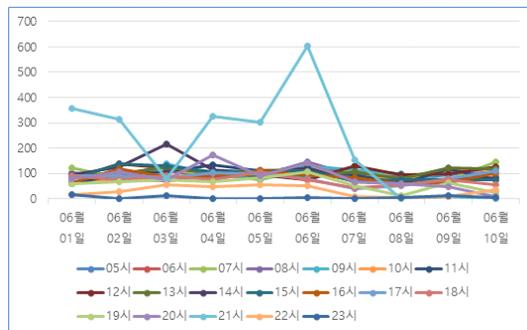
체크인카운터의 경우, 체크인카운터 방문자 수는 일반적인 시설 패턴과 유사하다. 아침 첫 운항 시간대, 점심 시간대, 이른 저녁 시간대에 방문자가 많음을 알 수 있다. 소요 시간과의 관계를 보면 방문자 수가 소요 시간에 영향을 미치지 않음을 알 수 있다.

6월 3일(월요일) 21시대가 가장 체류 시간이 매우 긴 것으로 나타났다. 그러나 이 기간에 운항에 특이 사항은 없었다. 운항 마감 시간과 가까운 시간대라 체크인카운터 시설 증설(3대)사업에 대한 인력 투입으로 예측된다.



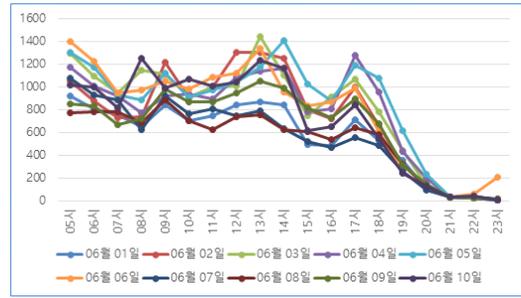
주: 세로축은 해당 시간대에 셀프체크인 카운터를 방문한 사람의 수를 나타냄(Wi-Fi AP에 접속자 수).

Fig. 7. Self check-in counter (number of visitors)



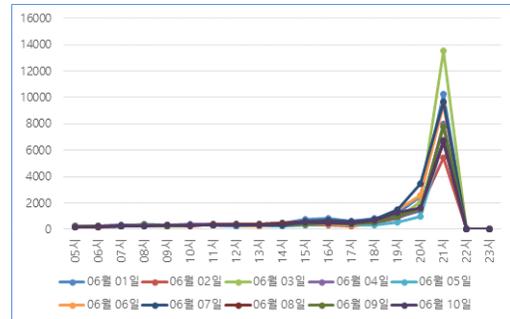
주: 세로축은 해당 시간대에 셀프체크인 카운터를 방문한 사람이 머문 시간(초)을 나타냄.

Fig. 8. Self check-in counter (staying time)



주: 세로축은 해당 시간대에 체크인카운터를 방문한 사람의 수를 나타냄(Wi-Fi AP에 접속자 수).

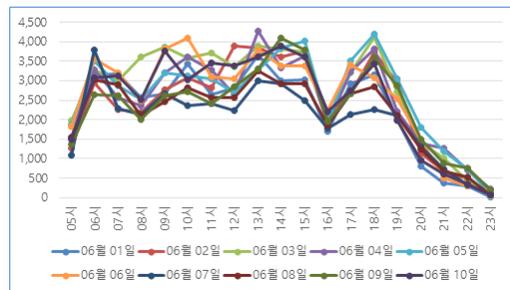
Fig. 9. Check-in counter (number of visitors)



주: 세로축은 해당 시간대에 체크인카운터를 방문한 사람이 머문 시간(초)을 나타냄.

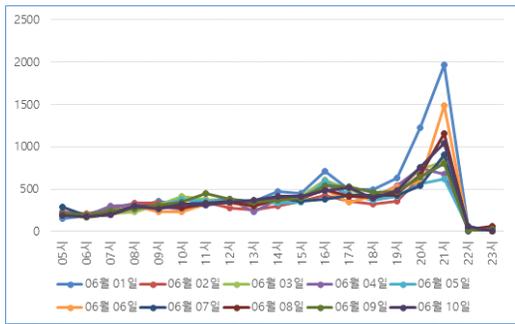
Fig. 10. Check-in counter (staying time)

탑승구의 경우, 방문자 수는 일반적인 시설 패턴과 유사하다. 아침 첫 운항 시간대, 점심 시간대, 이른 저녁 시간대에 방문자가 많았다. 타 시설과의 차이점은 유사 패턴과 더불어 운항 시간에 가장 많은 영향을 받음을 알 수 있다. 예를 들면 6월 6일 21시대 지연으로 인해 방문자 수가 감소한 것을 확인할 수 있다.



주: 세로축은 해당 시간대에 탑승구를 방문한 사람의 수를 나타냄(Wi-Fi AP에 접속자 수).

Fig. 11. Gates (number of visitors)



주: 세로축은 해당 시간대에 탑승구를 방문한 사람이 머문 시간(초)을 나타냄.

Fig. 12. Gates (staying time)

탑승구의 일자별 시간대별 소요 시간 역시 21시대가 가장 체류 시간이 긴 것으로 나타났다. 21시대 평균 소요 시간이 약 30분(1,800초)인 것으로 봤을 때 역시 상주 직원이 수집 자료에 포함된 것으로 추정된다.

탑승객의 수는 공항에서의 혼잡에 영향을 미칠 것이라는 전제하에 여러 가지 요인별로 연구가설을 설정하여 검증하였다. 가설검증을 위한 회귀분석(regression analysis)은 OLS(ordinary least square)방법을 이용하였다.

가설 1) 공항의 방문자 수( $V_t$ )는 공항 시설에서 방문자의 체류 시간에 영향을 준다.

회귀분석의 결과, 모형은 통계적으로 유의하게 나타났음을 볼 수 있다(Significance  $F=0.0021$ , d.f.=569,  $R^2=0.1443$ , 조정된  $R^2=0.1260$ ). 독립변수(방문자 수)는 부(-)의 영향을 미치는 것을 알 수 있다. ( $p$ -value=0.00219)

즉, 방문자의 수가 증가함에 따라 체류 시간은 줄어드는 것으로 나타났다. 이는 일반적인 추측과는 다소

Table 3. Regression analysis results for hypothesis 1

|      | 자유도      | 제곱합       | 제곱평균     | F 비    | 유의한 F  |
|------|----------|-----------|----------|--------|--------|
| 회귀   | 1        | 12485053  | 12485054 | 9.4667 | 0.0021 |
| 잔차   | 568      | 749096881 | 1318833  |        |        |
| 계    | 569      | 761581935 |          |        |        |
| 변수   | 계수       | 표준오차      | t통계량     | P-값    |        |
| Y절편  | 613.1314 | 74.1825   | 8.2651   | 0.0000 |        |
| 독립변수 | -0.13612 | 0.0442    | -3.076   | 0.0021 |        |

$y=613.1314-0.13612x$ .

y: 체류 시간, x: 방문자 수.

Table 4. Regression analysis results for hypothesis 2

|               | 자유도      | 제곱합       | 제곱평균     | F비    | 유의한 F |
|---------------|----------|-----------|----------|-------|-------|
| 회귀            | 1        | 10961140  | 10961140 | 7.881 | 0.005 |
| 잔차            | 538      | 748223411 | 1390750  |       |       |
| 계             | 539      | 759184551 |          |       |       |
|               | 계수       | 표준오차      | t통계량     | P-값   |       |
| Y 절편 (체류 시간)  | 629.6214 | 80.4430   | 7.8269   | 0.000 |       |
| 방문자 수 (1시간이내) | -0.0667  | 0.0237    | -2.807   | 0.005 |       |

$y = 629.6214 - 0.0667x$ .

y: 체류 시간, x: 방문자 수(1시간 이내 체류한 방문자 수).

다른 결과이기도 하다. 이러한 결과를 보다 논리적으로 타당하게 설명하기 위해 추가적인 가설을 검증하였다.

가설 2) 공항의 방문자 수(1시간 이내 체류한 방문자 수,  $V_{t-1}+V_t$ )는 공항 시설에서 방문자의 체류 시간에 영향을 준다.

연구가설 1과 마찬가지로 결괏값의 회귀 모형은 통계적으로 유의하게 나타났음을 볼 수 있다(Significance  $F=0.0051$ , d.f.=539,  $R^2=0.1639$ , 조정된  $R^2=0.1466$ ). 독립변수(방문자 수)는 역시 부(-)의 영향을 미치는 것을 알 수 있다( $p$ -value=0.005176).

방문자의 수를 1시간 이내 체류한 방문자의 수로 확대했을 때 결과는 동일하게 부(-)의 영향을 미치는 것으로 나타났다. 그러나, 계수의 크기는 줄어든 것을 알 수 있다(-0.1361에서 -0.0668로).

따라서 체류 시간에 영향을 미치는 방문자의 수는 포아송분포의 특성을 반영하여 지연(delay)이 존재한다는 것을 알 수 있고, 더 타당한 결과를 얻을 수 있었다.

이러한 기본모형의 결과를 바탕으로, 본 연구의 핵심 연구가설인 공항 시설별 체류 시간의 영향요인을 검증해 보았다.

가설 3) 공항의 시설(셀프체크인, 체크인카운터)별로 체류 시간에 차이가 있다.

가설 3의 실증결과에서는 연구모형에서 설정한 다수의 독립변수에 대해서 유의적인 결과가 나타났다(Significance  $F=0.0000$ , d.f.=570,  $R^2=0.1958$ , 조정된  $R^2=0.1893$ ).

다중회귀분석(multiple regression analysis)에서

Table 5. Regression analysis results for hypothesis 3

|                | 자유도      | 제곱합       | 제곱평균     | F 비    | 유의한 F |
|----------------|----------|-----------|----------|--------|-------|
| 회귀             | 4        | 72989937  | 18247484 | 19.99  | 0.000 |
| 잔차             | 566      | 688591998 | 1216594  |        |       |
| 계              | 570      | 761581935 |          |        |       |
| 변수             | 계수       | 표준 오차     | t 통계량    | P-값    |       |
| Y절편<br>(체류 시간) | 924.939  | 175.5854  | 5.27     | 0.0000 |       |
| 방문자 수          | -0.2293  | 0.0643    | -3.57    | 0.000  |       |
| 셀프체크인          | -672.265 | 158.5594  | -4.24    | 0.000  |       |
| 체크인카운터         | 93.9543  | 158.9116  | 0.59     | 0.554  |       |

$$y=924.939-0.229x_1-672.265x_2+93.954x_3$$

y: 체류 시간,  $x_1$ : 방문자 수,  $x_2$ : 셀프체크인,  $x_3$ : 체크인카운터

는 독립변수들간의 다중공선성(multicollinearity)의 문제를 검토해야 한다. 모델3의 독립변수는 방문자 수와 셀프체크인과 체크인카운터인데, 셀프체크인이 체크인카운터는 시설물을 나타내는 더미변수(dummy variables)로 부여되었기 때문에 다중회귀분석에서의 다중공선성의 문제가 존재하지 않는다.

셀프체크인 시설, 체크인카운터, 탑승구 시설의 방문자 수를 독립변수로 하고 회귀분석을 한 결과, 셀프체크인 시설과 방문자 수가 부(-)의 영향을 미치는 것으로 나타났다. 특히 셀프체크인은 계수의 크기가 방문자 수에 비해 매우 큰 것으로 나타나서 셀프체크인이 체류 시간의 축소에 영향을 미치는 것으로 나타났다. 이는 셀프체크인에서의 효율적인 흐름을 증명하고 있다.

모형의 설명력을 측정하는  $R^2$ 를 비교해 보면(모형의 독립변수의 수가 다르므로, 조정된  $R^2$ 로 비교할 때), 모형1(0.1260)과 모형2(0.1466)보다 모형3(0.1893)이 높아진 것을 알 수 있다. 즉, 모형의 설명력이 개선되었다는 것을 알 수 있다.

## VI. 결 론

본 연구의 목적은 여객 청사 내 여객 흐름에 따른 혼잡 시간과 및 주요 시설별 여객 체류 시간을 Wi-Fi 센서 수집을 통해 알아봄으로써 혼잡 시간대에 여객이 더 편안하게 공항을 이용할 수 있도록 의사결정에 도움을 주는 것이다. 이를 위해 먼저 국내 공항 중 주요 공항인 김포공항 내선을 중심으로 Wi-Fi 센서 데이터를 이용하여 그 유의성을 살펴보았다. 이를 위하여 주요 시설별 적정 혼잡도 면적을 산정하고, 그에 맞는

Wi-Fi 센서 반경 기준도 수립하였다. 그리고 그 기준을 토대로 시간대별로 주요 시설에서의 체류 시간을 해당 시설의 방문자 수와 해당 시설의 특성의 함수로 고려하여 회귀분석을 실시하였다. 본 연구의 주요 결과를 요약하면 다음과 같다.

첫째, 센서를 이용한 혼잡도 측정을 위해 현장 실사를 통해 여객 수를 카운팅한 결과, 약 95% 이상의 정확도를 보이는 것을 확인할 수 있었다. 또한, 센서에서 수집하는 자료가 개인 정보가 아닌 MAC 어드레스를 비식별화하여 유일한 Id 값으로 센서마다 실시간으로 많은 양의 자료를 수집한 후 처리함으로써 개인 정보 보호에 대한 문제 없이 여객 흐름을 분석할 수 있음을 알 수 있었다. 그러나 역설적으로 표현하면 개인에 대한 어떠한 정보, 예를 들면 성별, 연령대, 국적 등의 여객 특성을 분류할 수 있는 정보가 없으므로 단순히 여객 동선이나 체류 시간만을 분석할 수밖에 없다는 한계점을 알게 되었다.

둘째, 연구를 시작하기 전에는 당연히 여객이 많이 집중되는 시간대에는 여객의 체류 시간도 길어질 것으로 생각할 수 있다. 그러나 실제 연구를 진행한 결과, 탑승객 수가 많다고 해서 체류 시간이 길지는 않음을 알 수 있었다. 이렇게 된 원인으로 추측해 볼 수 있는 것은 상주 직원이 소유한 휴대전화기나 공항 청사 내 설치된 다양한 IT 장비에서 발생하는 신호까지 센싱했을 가능성이 크다. 특히나 이 부분은 현재 수집되는 센싱 데이터가 어느 정도 축적이 된다면 중복 자료에 대한 기준을 마련해서 제거하면 더욱 정교한 데이터를 얻을 수 있으리라 기대한다.

방문자 수가 증가할수록 체류 시간이 감소한다는 실증분석의 결과를 해석해 보면, 방문자 수가 많아지게 되면 이에 대한 공항의 대처 인력이나 설비도 증가하고, 또 탑승객들도 주변에 방문자 수가 많아지게 되면 자신의 업무만이 처리되면 다른 장소로 빠르게 이동하는 경향이 있다는 점을 설명하고 있다. 가설3의 분석 결과와 연계하여 연구 결과를 해석해 보면, 셀프체크인과 체크인카운터가 (-)와 (+)의 유의적 결과를 보였다는 것은 셀프체크인에서는 체류 시간이 감소하고, 체크인카운터에서는 체류 시간이 증가한다는 실증적인 결과를 보여준 것이다. 이는 공항에서의 IT 설비를 활용하여 공항 내 혼잡을 줄이고, 흐름을 효율적으로 유도하고자 했던 목표들이 실증분석에서 성취되었다는 것을 확인할 수 있었다.

셋째, 이 연구에서는 공항을 방문하는 사람들이 모

두 휴대전화기를 소지하는 것으로 가정하였고, 휴대전화기의 Wi-Fi 신호로 방문자의 수를 산정하였다. 여전히 휴대전화기를 소지하지 않고 방문하는 때도 있어서 이 부분에서는 이 연구에서 반영하지 못했다. 추후 연구에서는 방문자의 수와 움직임을 보다 정확하게 측정할 수 있는 도구를 찾아서 적용할 경우보다 정확한 연구 결과를 도출할 수 있을 것이다.

마지막으로, 공항을 운영하는 운영자의 노하우가 그 원인이 될 수 있다. 즉, 요일별, 시간대별 특성을 고려하여 혼잡 시간대에는 주요 시설에 인력을 추가 배치함으로써 오히려 혼잡 시간대에 상대적으로 더 많은 여객을 처리함으로써 체류 인원 대비 시설별 체류 시간이 더 짧을 수 있다.

이 부분은 공항 운영의 주요 정책적 결정과의 관계를 통해 추후 연구가 필요한 부분이다. 즉, 탑승객 수가 많으면 체류 시간이 길어질 것으로 생각할 수 있으나, 공항의 시설별로 탑승객 수에 따른 적절한 조치는 체류 시간을 축소할 수 있다는 점을 보여줌으로써 공항 운영의 특성에 따른 혼잡도 조절의 가능성을 보여 주었다는 것에 의의가 있다.

## Ⅶ. 연구의 한계 및 제언

본 연구에서는 항공교통이용자의 특성과 프로세서가 고려되지 않았다. 이 연구에서는 공항의 설비를 기준으로 측정하였다. 개인정보보호의 규정에 따라 공항 설비에서 측정한 정보로 이용자의 특성은 분석할 수 없었다. 이 연구에서는 공항의 주요 설비에서 측정할 경우 더 많은 실측 정보를 확보할 수 있을 것이라는 점에서 이 방법론을 채택하였다. 항공 이용자의 특성을 반영한 추가적인 연구는 후속 연구의 대상이 될 것이다.

본 연구는 Wi-Fi 센서가 일정 간격을 두고 자료를 수신함에 따른 동일 자료의 이중 측정, 상주 직원이나 각종 공사에 따른 일시적인 장기 체류 인원, 공항 내 장비에서 발생하는 센서 자료까지 측정되는 부분으로 더 정확한 측정 도구가 필요할 것으로 판단된다. 요즘 휴대전화기의 Mac 어드레스가 고정 주소가 아닌 가변 주소를 사용함에 따라 동일 휴대전화기를 이중으로 인지하는 때도 발생하였다. 이 부분은 센서 수집 기술이 계속 발전함에 따라 추후 연구에는 보완이 될 것으로 기대한다.

본 연구는 이용객의 데이터를 분석하고자 하였는데, 이용객이 아닌 사람의 데이터를 완전하게 배제하지 못

했다. 이 부분은 분석 결과에 대한 신뢰성을 저해할 수 있다. 기술적으로는 이러한 데이터를 배제할 수 있으나, 이 경우 개인정보침해의 소지가 있어서 현재로서는 어려운 측면이 있었다. 기존 연구들이 설문 응답을 통한 간접적인 조사에 머물렀지만, 본 연구는 더 많은 양의 객관적인 자료를 활용하였다는 점에서는 나름의 차별성을 갖는다. 그런데도, 완벽한 형태의 자료수집이 이루어지지 못했다는 점에서는 연구의 한계로 남는다. 추후 연구에서는 이런 부분을 극복할 수 있도록 이 연구가 탐색적 연구로서의 기초가 될 수 있기를 기대한다.

모든 공항의 보편적 문제는 이용 시간이 특정 시간대로 집중된다는 점이다. 공항 운영자는 특정 시간 침투시간대 이용자 동선과 항공편 배치가 주요 고려사항이다. 공항에서는 이용자가 선호하는 특정 시간대에 비행편이 집중되는 현상이 발생한다. 공항은 요소별 지연 시간을 해소하기 위해 이용자의 동선과 편의성을 위한 위치 배정을 고려하고 있다. 본 연구에서는 이러한 부분을 반영하지 못했다. 추후 연구에서는 이러한 공항의 고려사항을 고려한 연구가 진행될 것으로 기대한다.

## References

1. Schwab, Klaus, "The Fourth Industrial Revolution," Currency, 2017.
2. Park, C.-K., "A study on the passenger circulation system of the airport passenger terminal", Korean Institute of Interior Design Journal. 19(3), 2010, pp. 260-268.
3. Bae, Y., "A study on the improvement of size-estimation airport passenger terminal using cause and effect diagram. department of construction management", Graduate School of Engineering of Hanyang University, 2018.
4. Kang, Y. J., "A study for enhancing performance measure of a passenger terminal simulation model", Master Thesis, Graduate School of Korea Aerospace University, 2017.
5. Seo, H. J., "A study on the customer behavioral analysis using big data of distribution industry", Department of Applied Statistics, Graduate School of Gachon University, 2017.
6. An, H., "A study on the sentiment analysis in

- fashion design using big data - Focused on text mining and semantic network analysis”, Department of Clothing & Textiles, The Graduate School, Ewha Womans University, 2017.
7. Turban, E., Pollard, C., Wood, G., “Information Technology for Management”, 11th edition, Wiley Custom. ISBN: 978-1-119-92381-7, 2018.
  8. Kim, J.-S., “Big data analysis technologies and practical examples”, The Journal of the Korea Contents Association, 12(1), 2014, pp. 14-20.
  9. Park, R. U., “A plan for activation of local tourism through big data with focus on the Jeollabuk-do”, Department of Integrated Bio-Resource Science, Graduate School of Jeonju University, 2017.
  10. Kim, J. S., “A big data based spatio-temporal sensor data processing system”, Doctoral Thesis, Graduate School of Konkuk University, 2017.
  11. Lee, Y. J., “A study on airport parking lot scale for international airports - Focused on Incheon international airport”, Department of Business Administration, Graduate School of Aviation & Management of Korea Aerospace University, 2013.
  12. Moon, D. H., “An establishment of big data at a corporation and a study on the application”, Graduate School of Business of Gyeong Sang National University, 2016.
  13. Choi, M. K., “A study on the indoor evacuation path model using space big data of IoT sensor”, Department of Geography, Graduate School of Kyung Hee University, 2018.
  14. Kim, J. S., “Subway congestion prediction and recommendation system using big data analysis”, Journal of Digital Convergence, 14(11), 2016. pp. 289-295.
  15. Lee, J. G., “A study on improvement of local traffic information providing using big data”, Department of Software Engineering, Graduate School of Information Science, Soongsil University, 2017.
  16. Kim, H., “Development of a congestion index for expressway service areas using floating population big data”, Graduate School of Transport & Intelligent Transport Systems of Ajou University, 2017.
  17. Han, Y. H., “Development of spatiotemporal congestion recognition index based on big traffic data”, Doctoral Thesis, Graduate School of City University of Seoul, 2018.
  18. Min, Y. S., “A study on the efficiency of the national suicide prevention project using big data analysis”, Department of Forensic Science of Sungkyunkwan University, 2019.
  19. Yoon, J. H., “A study on applying big data analytics in the electric power industry”, Graduate School of Information Science, Yonsei University, 2017.
  20. Lee, S.-J., “The determinants of the implementation of the big data in traffic fields: perspectives of data quality management”, Department of Management Information System, Graduate School, Keimyung University, 2018.
  21. Ministry of Construction and Transport, “The Second Airport Development Long-term Master Plan”, 1999.
  22. Ministry of Land, Infrastructure and Transport, “The Third Airport Development Long-term Master Plan. Chapter 6. Design and Capacity Estimate of Airport Facility”, 2005.
  23. Nepal, R., Al Irsyad, M. I., Nepal, S. K., “Tourist arrivals, energy consumption and pollutant emissions in a developing economy-implications for sustainable tourism”, Tourism Management, 72, 2019, pp. 145-154.