

모바일 통신 빅데이터 기반 항공교통이용자 O/D 추출 알고리즘 연구

Algorithm Development for Extract O/D of Air Passenger via Mobile Telecommunication Bigdata

조범철 · 권기훈[†]

한국교통연구원 교통빅데이터연구본부

요약

현행 항공교통이용자 분석은 주로 통계적인 분석이 주류를 이루고 있으나, 이동경로, 지역별 이용자 수, 공항접근 소요시간 등 세부적인 사항에 대한 분석이 어렵다는 한계가 있다. 한편 빅데이터 기술 발전과 데이터3법 개정에 따라 빅데이터 기반 교통분석이 활성화되고 있으며, 모바일 통신 데이터는 휴대전화 단말기의 위치를 상세하게 파악할 수 있어 교통분석을 위한 좋은 분석자료가 될 수 있다. 이에 본 연구에서는 기존 항공교통이용자 분석방법의 한계를 극복하기 위해 이동경로 전체를 분석할 수 있는 모바일 통신 데이터를 기반의 교통이용자 O/D(Origin/Destination) 추출 알고리즘을 제시한다.

본 연구에서 제시하는 알고리즘은 각 공항에 공항신호탐지 구역 설정하고, 해당 구역의 기지국 접속이력을 토대로 항공교통이용자를 추출하고 해당 이용자의 출발지-도착지 경로상의 기지국 접속 데이터를 토대로 이동경로를 추정하는 것이다. 본 연구에서는 2019년 1~12월의 기간을 대상으로 모든 국내 공항에 대하여 국내/국제선 이용자에 대해 O/D를 추출하였다. 또한 추출된 데이터의 검증에 위해 모바일 통신데이터 기반 항공교통 이용자 O/D 데이터와 항공통계데이터에 대해 상관성 분석을 수행하였다. 이를 통해 총량에는 차이가 있으나(국내선 4.1, 국제선 4.6) 상관성 0.99로 상관성이 높아 활용 가능할 것으로 판단되었다.

본 연구에서 제시한 알고리즘은 기존과 다르게 항공교통이용자의 이동행태, 지역별/연령별 비율 등 폭넓고 상세한 분석을 가능하게 하며, 향후 공항관련 정책 마련이나 지역별 시장분석 등 다양한 분야에서 활용할 수 있다.

■ 중심어 : 빅데이터, 모바일 통신, 트립체인, 항공교통이용자, 공항수요

Abstract

Current analysis of air passengers mainly relies on statistical methods, but there are limitations in analyzing detailed aspects such as travel routes, number of regional passengers and airport access times. However, with the advancement of big data technology and revised three data acts, big data-based transportation analysis has become more active. Mobile communication data, which can precisely track the location of mobile phone terminals, can serve as valuable analytical data for transportation analysis.

In this paper, we propose a air passenger Origin/Destination (O/D) extraction algorithm based on mobile communication data that overcomes the limitations of existing air transportation user analysis methods. The algorithm involves setting airport signal detection zones at each airport and extracting air passenger based on their base station connection history

2023년 10월 04일 접수; 2023년 12월 07일 수정본 접수; 2023년 12월 08일 게재 확정.

* 본 연구는 국토교통부의 지원으로 수행되었음(발간등록번호 11-1613000-003229-01)

† 교신저자 (kwonkh@koti.re.kr)

within these zones. By analyzing the base station connection data along the passenger's origin-destination paths, we estimate the entire travel route.

For this paper, we extracted O/D information for both domestic and international air passengers at all domestic airports from January 2019 to December 2020. To compensate for errors caused by mobile communication service provider market shares, we applied an adjustment to correct the travel volume at a nationwide citizen level. Furthermore correlation analysis was performed on O/D data and aviation statistics data for air traffic users based on mobile communication data to verify the extracted data. Through this, there is a difference in the total amount (4.1 for domestic and 4.6 for international), but the correlation is high at 0.99, which is judged to be useful.

The proposed algorithm in this paper enables a comprehensive and detailed analysis of air transportation users' travel behavior, regional/age group ratios, and can be utilized in various fields such as formulating airport-related policies and conducting regional market analysis.

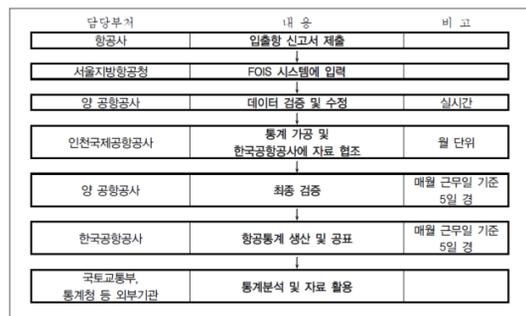
■ Keyword : Bigdata, Mobile Telecommunication, Trip Chain, Air Passenger, Airport Demand

I. 서론

최근 빅데이터 기반 기술의 발전에 따라 빅데이터를 기반으로 한 다양한 교통분석방안 연구가 진행되고 있다. 특히 개정된 데이터 3법(개인정보보호법, 정보통신망법, 신용정보법)을 통해 다양한 분야의 빅데이터를 기반으로 한 교통분석의 가능성이 활짝 열렸다고 볼 수 있다.

이러한 빅데이터 기반 분석방법론은 이미 교통수요예측, 지역경제활동 모니터링, 교통정책 수립 등 다양한 분야에 활용되고 있다. 이는 항공분야에서도 마찬가지이다.

항공분야에서는 항공수송에 대한 변동추이 파악을 위하여 항공통계를 작성하고, 이를 공항 건설, 운영, 관리 및 항공정책 수립 등에 활용하고 있다. 항공통계는 통계법 제18조에 의해 승인된 일반통계로, 항공사, 서울지방항공청, 공항공사 등에서 수집한 데이터를 토대로 집계되는 자료이다. 주요 항공관련 통계는 항공사, 서울지방항공청, 한국공항공사, 인천국제공항공사 등 4개 기관이 직간접적으로 관여하며 상호 자료를 수집 및 집계한다[1].



〈그림 1〉 국내 공항 통계정보 수집 프로세스

서울지방항공청은 FOIS(Flight Operation and Information System, 비행 및 운항정보시스템), 한국공항공사는 iFIS(Integrated Flight Information System, 통합운항정보시스템), 인천국제공항공사는 IIS(Integrated Information System, 통합정보시스템)를 각각 운영하고 있으며, 해당 시스템을 통해 자동으로 수집된 항공사별 수송실적 자료들은 통계담당자의 검증을 거쳐 통계자료로 생산된다. 또한 해당 항공통계는 공공데이터의 제공 및 이용 활성화에 관한 법률 제14839호(2017.7.26.)에 따라 여러 곳에서 제공되고 있다. 해외의 경우, 우리나라와 달리 관련자료 수집이 제한되어 있으며, ACI(Airports Council International, 국제공항협의회), OAG(Official Airline Guides) 등에

서 일부 실적데이터를 구매할 수 있다.

그러나 이러한 통계데이터는 이용자 총량이나 각 기간별 이용자 변화 추이 등을 분석할 수는 있으나 세부적인 사항의 분석에는 한계가 있다. 특히 공항접근구역(Catchment Area)의 경우, 각 공항에 대하여 어느 지역의 이용자가 방문하는지, 공항까지 이동할 때 어느 정도의 시간을 소모하는지 등에 대하여 분석하기 위한 것으로, 이에 대한 분석을 위해서는 각 항공교통이용자의 최초 출발지, 최종목적지의 확인과 함께 행정동, 시간단위의 데이터 수집이 필요하다. 그러나 기존의 통계데이터 중심의 분석으로는 이에 대한 분석이 어려우며 기존과는 다른 방식의 데이터가 요구된다.

따라서 본 연구에서는 전국, 전 인구를 대상으로 항공교통이용자를 분석하기 위한 기초자료를 만들기 위하여 모바일 통신 기지국 데이터를 활용하여 항공교통이용자의 O/D를 추출하기 위한 알고리즘의 연구를 수행하였다. 또한 해당 알고리즘을 통해 도출된 데이터의 검증은 위해 코로나19 팬데믹의 영향이 없는 2019년을 대상으로 데이터 추출 및 검증을 수행하였다.

II. 선행연구

항공분야에서는 항공교통이용자의 분석을 위해 주로 통계데이터를 토대로 회귀분석 모형 등 다양한 방법을 이용해 분석을 수행해왔다. 예를 들어 수요예측은 크게 정성적 분석, 정량적 기법, 결합기법 등을 활용하게 된다.[2] 항공분야에서는 주로 단기예측의 경우, decomposition models, exponential smoothing, ARIMA 모형을, 중장기예측은 linear regression, log-linear, linear-log, log-log, trend projection, KENZA model 등을 활용하고 있으며 이는 항공통계데이터를 활용하여 수행된다.[2]

또한 공항별 수요 예측이나 지역별 공항이용

수요 예측을 위한 모형의 연구가 진행되어 왔다. 아주대학교의 경우, 집계 관측 자료를 이용한 국제선 여객 공항선택 모형을 연구하였다.[4] 해당 연구는 공항 접근통행시간과 비용, 이용대상 항공노선 운항빈도를 설명변수로 설정하여 국제선 공항 선택 형태를 분석하고자 한 연구로, 관측하기 어려운 지방의 고유특성을 반영하여 패널회귀방법을 적용한 영남권 국제선 항공수요 분석을 수행하였다. 이를 통해 고정효과 모형(Fixed-effects model)이 가장 적합한 모형으로 밝히고 있다.

박재성 외 3인의 연구(2016)에서는 항공수요에 대한 계량경제모형 개발을 위해 시계열 자료의 특성을 검증하기 위한 단위근 검정을 수행하였으며, 변수들 간의 장기 균형관계 분석을 위해 공적분 검정을 수행하였다. 모형 적용을 위해 제주공항 국내선을 분석하였으며, GDP와 항공요금지수, 그리고 제주 방문 관광객 수 등 통계자료를 활용하여 수요 모형의 설명변수로 사용하였다[5]. 김제철 외 2인의 연구[6]에서는, 국내공항의 중단기적 국제선 수요예측 시 단변량 시계열 모델인 ARIMA 모형에서 개입효과를 적용한 ARIMA-Intervention 모형을 활용하여 bottom-up 방식의 국제선 항공여객 수요에 대한 실증 예측분석 결과를 제시하였다. 해당 분석에서 메르스 사태, 사드배치에 따른 한국 여행상품 판매 증지를 고려하였으며, 2018년 상반기 실적 대비 약 1.78%의 오차가 발생하여 안정적인 예측이 가능함을 보였다. 김영록의 연구에서는 국제선 여객수요 및 화물수요 관측을 위해 OLS 다중회귀 분석을 수행하였다[7]. 이 과정에서 한국공항공사의 항공통계, 한국은행 경제통계시스템 등 국내 기관에서 제시하는 자료를 활용하여 각 분기별 자료에 대해 OLS 다중회귀분석을 수행하였다.

이와 같이 그 동안의 항공수요 예측은 많은 경우, ARIMA 모형 등의 시계열 분석 방법이나 linear regression 등의 회귀분석 모형을 통해 이

루어져 왔다. 또한 분석에 활용되는 자료도 항공통계, 공항집계자료 등 통계적인 자료를 주로 사용하고 있다.

그러나 이러한 선행 연구들은 단순히 통계데이터를 기반으로 한 통계적 분석을 중심으로 하고 있다. 이는 항공교통이용자의 이동경로, 지역별 항공교통이용자 수, 항공교통이용자의 공항 접근 소요시간 등 항공교통이용자의 통행행태와 같은 세부사항을 분석하는 데는 한계점이 있다. 최근 미래교통수요의 분석을 위해 이동목적, 이동수단 및 인구사회구조 변화를 고려한 통행행태를 분석하는 연구[8]도 이루어지고 있으며, 지역별 수요, 접근교통 관련 분석 및 정책 마련 등을 위한 기초자료로서 보다 공항이용자의 통행행태를 분석할 필요가 있다.

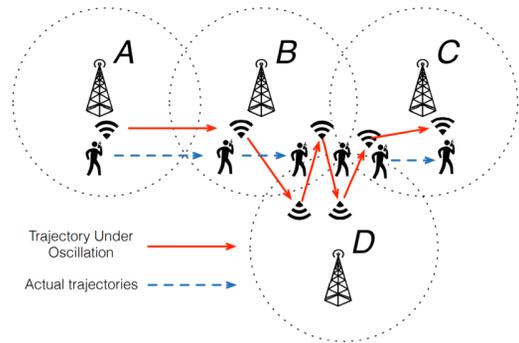
이에 따라 기존의 수요예측 방법과 달리 항공교통이용자의 이동경로를 도출하고 이를 바탕으로 항공수요 및 항공교통이용자의 이동행태에 대한 근본적인 분석을 위한 새로운 분석방법, 즉 모바일 통신데이터를 기반으로 한 항공교통이용자 분석 방안의 연구가 필요하다.

III. 모바일 통신 빅데이터 기반 O/D 추출 알고리즘 연구

3.1 모바일 통신 빅데이터

모바일 통신 빅데이터를 간략히 설명하면, 모바일 통신 사용 시에 발생하는 신호, 기록 데이터를 기반으로 하는 빅데이터이다. 해당 데이터는 CDR(Call Detail Record, 이동통신 단말기 이용시 발생 데이터), Signaling(통화단절 최소화를 위해 휴대전화-기지국 간 지속적 송수신 내역 기록)이 있으며, 각 기지국에 기록된 데이터를 활용하여 해당 휴대전화 단말기가 언제, 어느 지역에 있었는지 파악할 수 있다.

모바일 통신 빅데이터는 해당 휴대전화 단말



〈그림 2〉 모바일 통신 셀 구성 및 핸드오버 데이터 수집

자료: Qiao et al., “A Mobility Analytical Framework for Big Mobile Data in Densely Populated Area”, IEEE Transactions on Vehicular Technology, 66(2), 2017, pp. 1443-1455.

기를 보유한 이용자의 위치를 365일, 24시간에 걸쳐 파악이 가능하다는 특징이 있다. 이는 대부분의 휴대전화 이용자가 단말기를 항상 소지하고 있으며, 언제나 전원이 켜진 상태로 유지하기 때문이다. 또한 접속 이력이 남은 각 기지국의 위치에 따라 행정구역을 매칭하고, 각 이용자의 기지국 별 체류시간을 분석하여 출발지, 도착지, 각 체류지역 분류(거주지, 직장 등) 등을 분석할 수 있다. 이러한 특성은 통계결과에 의존하는 기존의 방법과 다르게 공항 또는 항공교통이용자의 1일간의 이동경로 자체를 분석할 수 있게 한다.

따라서 모바일 통신 빅데이터는 단순히 공항 방문자, 항공교통이용자 숫자만 분석하는 것을 넘어 항공교통이용자가 공항에 오기 이전에 어디서 출발했는지, 항공교통을 이용한 뒤 어디로 이동하였는지와 같은 이동경로, 이동패턴, 이동목적 등의 분석에 활용할 수 있다. 또한 각 공항별 이용자의 출/도착지 분포 분석으로 각 공항의 접근 영향권 분석에 활용할 수 있다. 이에 더해 개인정보보호법 등 국내 규정에 따라 분석하기 어려운 사항에 대한 분석도 가능하다. 예를 들어 항공교통이용자의 공항체류시간을 분석할

시, 기존의 CCTV 등 공항 터미널 내 스캔장비 활용은 개인정보보호법에 따라 활용이 불가하나 모바일 통신 빅데이터는 공항구역 내 기지국 접속시간을 토대로 공항도착-여객기 이륙 시까지의 체류시간을 추정하는데 활용할 수 있다.

3.2 모바일 통신 빅데이터 기반 항공교통 이용자 O/D 추출 알고리즘

본 연구에서 제시하는 모바일 통신 빅데이터 기반 항공교통이용자 O/D 추출을 간략히 설명하자면 모바일 통신기지국의 로그데이터를 토대로 항공교통이용자를 추출하고, 해당 항공교통이용자들의 세부 이동경로, 중간체류지 등의 데이터를 추출하는 것이다.

각 공항에는 해당 공항의 이용에 따른 운송수요 발생 지리적 지역인 공항접근영향권(Airport Catchment Area)이 존재한다. 이러한 공항접근영향권과 공항별 운송수요 등의 상세분석과 예측을 위하여 항공교통이용자의 상세한 분석이 필요하며, 이를 위해 항공교통이용자가 공항을 이용하기까지, 그리고 공항을 이용한 이후 목적지까지의 상세한 경로를 추출하고 분석할 필요가 있다.

이에 따라 항공교통이용자의 출/도착지역, 이동행태, 공항까지 이동하는데 소요되는 시간, 지역별 공항이용자 수 등 상세한 분석을 위해서는 이동경로 및 이동행태의 분석이 필요하다. 따라서 항공교통이용자의 O/D (Origin/Destination) 데이터는 출발지, 도착지, 이용공항, 중간경유지

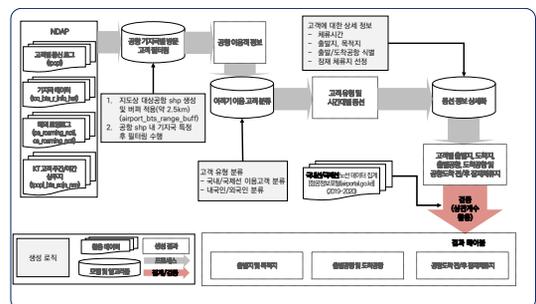
등을 포함하고 있어 이동경로, 이동행태 등을 분석할 수 있는 기초자료가 될 수 있다.

이를 위하여 본 연구에서 제시하는 알고리즘은 항공교통이용자의 트립체인 도출을 목표로 항공교통이용자가 공항을 거쳐 목적지까지 이동하는 과정 중 통과한 지역의 통신기지국에 기록된 접속기록을 토대로 최초 출발지 및 최종 목적지, 중간 체류지를 추정하는 알고리즘이다. 여기에 모바일 통신사에서 보유한 회원정보와 결합하여 연령대, 내외국인 구분 등 보다 상세한 데이터를 추출할 수도 있다.

모바일 데이터 기반 항공교통이용자 O/D 추출은 크게 4단계로 이루어진다. 첫 번째는 공항 기지국 방문자 필터링으로, 공항에 설치된 모바일 통신기지국에 접속한 이력을 토대로 공항방문자를 추출하고, 타 공항 기지국의 접속기록을 비교하여 항공교통이용자를 추출한다. 두 번째는 여객기 이용 고객 분류로, 출/도착지 공항의 위치를 파악하여 국내선/국제선 구분 등을 수행한다. 세 번째는 동선정보 상세화로 항공교통이용자의 출발지, 목적지, 잠재체류지 등을 추정하여 트립체인을 추출하는 단계이다. 마지막으로 네 번째는 국내선/국제선 통계 데이터를 토대로 추출된 데이터의 정확성을 검증하는 것이다.



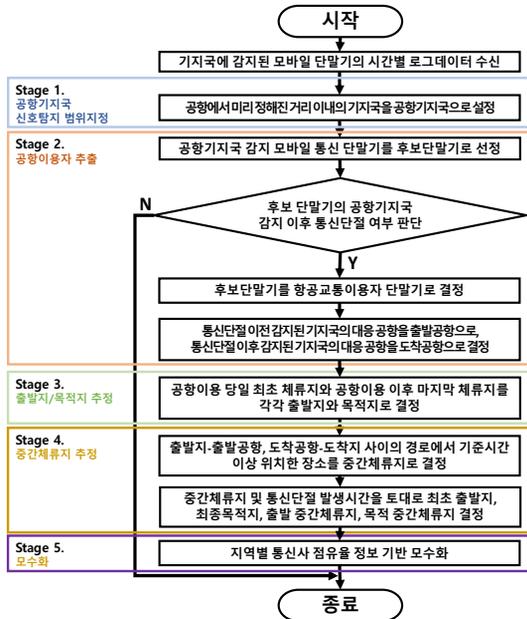
〈그림 3〉 모바일 데이터 기반 O/D 추정 개념도



〈그림 4〉 모바일 데이터 기반 항공교통이용자 O/D 추출 흐름도

본 연구에서 제시하는 항공교통이용자 O/D 추출 알고리즘은 최초 출발지, 최종목적지, 이용

공항, 중간체류지를 추출할 수 있도록 구성되었다. 이를 그림으로 간략히 표현하면 다음과 같다.



〈그림 5〉 항공교통이용자 O/D 추출 알고리즘 흐름도

또한 해당 알고리즘에 적용된 주요 기준은 아래와 같다.

공항이용자 구분 - 공항 범위에 속한 기지국 범위 확인 - 데이터 생성방위(2019~2020년내 공항기지국 로그가 1회 이상인 경우만 추출)	국제선/국내선 이용객 구분 - 개인 로그 중 하루 내 공항 1곳만 방문 → 국제선 - 개인 로그 중 하루 내 공항 2곳 이상 방문 → 국내선
출발공항/도착공항 구분 - 국제선 이용객의 경우 로밍 로그를 확인하여 출발 및 도착공항 구분 - 국내선 이용객의 경우 최초 방문한 공항은 출발공항 / 마지막 방문한 공항은 도착공항으로 처리	출발지/목적지 구분 - 공항 이용 전 3시간 이상 체류지 중 최초 행정동 → 출발지 - 공항 이용 후 3시간 이상 체류지 중 마지막 행정동 → 목적지 (출발/목적지가 국외인 경우 해당 국가명으로 구분)
출발지/목적지/중간체류지 측정 부위 - 해당 개인로그의 주안 및 야간 상주기록을 활용 - 주간 상주지역의 경우 직장으로, 야간 상주지역의 경우 집으로, 그 외의 경우는 기타로 분류 (poppl, bts, sgin, mmm 활용)	출발/도착/중간체류지 구분 - 출발 : 출발공항 방문 전 20분 이상 체류지 - 도착 : 도착공항 방문 후 20분 이상 체류지

〈그림 6〉 항공교통이용자 O/D 추정 주요 기준

3.2.1 공항기지국 신호탑지 범위 지정

항공교통이용자 O/D 추출에 앞서 우선적으로 공항방문자를 분류할 방법이 필요하다.

국내의 공항에는 이동통신 서비스 제공을 위한 모바일 통신기지국이 설치되어 접속기록 분석을 통해 공항방문자를 추정할 수 있다. 이를

위해 각 공항에 대하여 공항신호탑지역역을 지정하고, 해당 구역 내의 기지국을 분석대상으로 삼는다. 일반적으로 4G/5G 모바일 통신은 각 기지국별로 0.9Km의 관할구역을 가지고, 공항은 특성상 넓은 보호구역(Air Side)을 가지고 있어, 공항방문자 체류구역과 주변지역 사이에 상당한 거리가 있다.

이에 따라 본 연구에서는 전국 15개 공항에 대하여 공항 시설물 기준으로 2.5km의 버퍼를 부여하여 직사각형 형태로 공항 기지국 신호탑지 범위를 지정하였다.



〈그림 7〉 공항기지국 신호탑지 범위 지정 예시(인천국제공항)

3.2.2 항공교통이용자 추출

항공교통이용자의 추출은 공항 신호탑지 범위 내의 기지국의 접속기록을 이용하여 이루어진다.

항공기 이용시 발생하는 접속행태(출발공항 접속, 비행 중 단절, 도착공항 접속)를 반영하여 1일간 서로 다른 공항에 접속 이력이 발생하고, 그 사이에 타 지역 접속이력이 없다면 해당 단말기의 사용자는 항공교통을 이용했다고 판단할 수 있다. 이에 따라 국내선 항공교통이용자를 추출할 시, 서로 다른 공항을 방문한 휴대전화는 먼저 추출하고, 두 공항 지역의 기지국 접속시간 사이에 접속이력이 없는 사용자를 다시 추출하여 항공교통이용자를 추출할 수 있다. 이는 공항방문 당일 복수의 공항구역 접속이력이 있고, 그 사이에 접속이력이 없는 휴대전화만

추출하여 항공교통을 이용하지 않은 단순 공항 방문객(공항 상주인력, 환승객 등)을 제외할 수 있다.

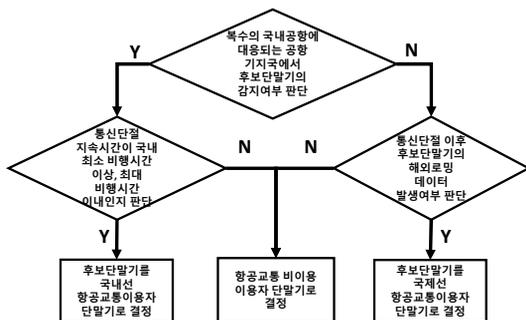
상기 추출 방법은 그림 5에 묘사된 Stage 2. 항공교통이용자 추출과 같다.

	출발공항 기지국	공항 간 중간경로 기지국	도착공항 기지국 접속
항공 이용자	로그 있음	로그 없음	로그 있음
출발공항 방문자	로그 있음	로그 없음	로그 있음
도착공항 방문자	로그 없음	로그 없음	로그 있음
육상교통 이용자	로그 있음	로그 있음	로그 있음

〈그림 8〉 항공교통이용자 추정 기준

또한 추출 정확도 향상을 위하여 공항이용 당일 에 한하여 항공교통이용자를 추출하며, 공항 간 이동시간은 국내선 최대 비행시간 이내로 설정하였다. 현재 국내선에서 가장 긴 비행시간을 가진 노선은 서울-제주, 양양-제주노선 등이며, 해당 노선의 비행시간은 1시간 10분 가량이다. 그리고 국내공항은 비행소음으로 인하여 커푸타임(Curfew Time)을 적용하고 있으며, 제주공항의 경우, 23~06시 사이의 운항을 제한하고 있다. 이에 따라 현행 비행스케줄은 당일 이내에 출/도착이 이루어지게 된다.

한편, 국제선 이용의 경우, 해외에 위치한 공항에 국내 통신사의 이동통신 서비스가 제공되지 않아 국내 모바일 통신 기지국 데이터로 식



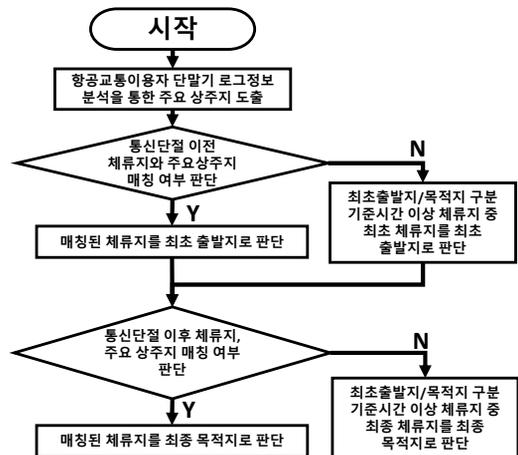
〈그림 9〉 국내선, 국제선 구분 알고리즘 흐름도

별하는데 한계가 있다. 이에 따라 국제선 이용자는 국내통신사와 연계한 해외통신사(NTT 등)에서 확인되는 로밍기록을 이용하여 해외 출/도착 공항을 분석하게 된다. 특히 국제선은 장거리 노선이 많아 출발 당일이 아닌 다음날에 도착하는 경우가 많아, 공항이용 전날 및 다음날의 해외로밍 기록을 추출하고, 공항이용 당일 국제공항 1개소만 방문한 사용자를 국제선 이용자 로 추정한다.

3.2.3 출발지 및 목적지 추정

항공교통이용자의 O/D를 추출하기 위해서는 최초 출발지와 최종 목적지를 도출해야 한다. 항공교통이용자의 이동행태를 고려하여, 출발지는 공항이용 당일 일정시간 이상 체류한 지역의 첫 번째 체류지로 지정하고, 목적지는 공항이용 후 일정기간 체류한 지역 중 당일의 마지막 체류지로 지정한다. 단, 공항도착 이후 새벽 시간까지 이동하는 경우를 고려하여 미확인 시 익일 오전 8시까지의 체류지를 확인하여 목적지를 찾게 된다.

국내선 이용 시에는 출/도착지를 모두 파악할 수 있으나, 국제선 이용 시는 해외 통신사 기록의 확보 및 분석이 어려워 출/도착지 추정에 한



〈그림 10〉 출발지, 목적지 도출 알고리즘 흐름도

계가 있다. 따라서 해외 로밍기록을 토대로 방문한 해외공항의 국가 및 공항명을 추정해야 한다. 이를 위하여 해외에 위치한 출발공항과 국가는 공항이용 당일 또는 전일의 해외로밍기록이 발생하는 국제공항 및 국가로 정의한다. 도착공항도 공항이용 당일 또는 다음날 해외로밍기록이 발생하는 국제공항 및 국가로 정의한다. 이는 유럽과 같이 국경이동이 쉬운 국가에서 목적지에 도달하기 위해 타국의 공항을 이용하는 경우를 고려한 것이다.

또한 출발지/목적지에서는 장시간 체류할 가능성이 높아 중간 체류지와와의 구분을 위한 체류시간 기준 마련이 필요하며, 확인비율이 가장 높은 시간기준인 3시간으로 지정하였다.

체류시간 기준	출발/목적지 확인 비율	
	출발지	목적지
3시간 ↑	96.12%	95.56%
4시간 ↑	93.19%	51.29%
5시간 ↑	88.45%	39.41%

〈그림 11〉 국내선 출발지 및 목적지 탐색 체류시간 EDA (Exploratory Data Analysis, 탐색적 데이터 분석) 결과

〈표 1〉 국제선 출/도착지 추정 기준

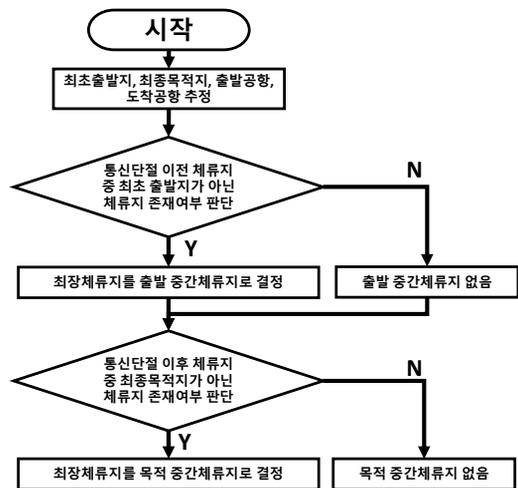
대상	상세
출발공항 (국외)	- (공항)공항이용 당일 방문한 국제공항 - (국가)공항이용 전날 해외로밍 확인되는 국가
도착공항 (국외)	- (공항)공항이용 당일 방문한 국제공항 - (국가)공항이용 다음날 해외로밍 확인되는 국가
출발지	- 공항방문 이전 일정시간 이상 체류한 지역 중 공항이용 당일 첫 번째 체류지
목적지	- 공항이용 후 일정시간 체류한 행정동 중 당일 마지막 체류지
체류시간 기준	- 3시간 이상

3.2.4 중간체류지 추정

이동경로의 상세한 분석을 위해서는 환승지

점 등 이동경로를 추정하기 위하여 중간지점을 추정하기 위한 방안의 마련이 필요하다.

본 연구에서는 공항과 출발지/목적지 사이에서 일정시간 이상 체류한 지역을 잠재체류지로 정의하고 이 중 가장 긴 시간을 체류한 장소를 중간체류지로 지정한다. 잠재체류지는 출발/목적 잠재체류지 2개로 구분되며, 이는 각각 출발지-출발공항, 도착공항-목적지 사이에 20분 이상 체류한 지역을 모두 잠재체류지로 지정한다. 이후 해당 구간에서 가장 긴 시간을 체류한 1개소를 중간체류지로 지정한다. 이러한 중간 체류지는 출발지-출발공항 사이의 출발 중간 체류지와 도착공항-목적지 사이의 목적 중간체류지로 나뉜다. 중간 체류지는 출발지/목적지-공항 사이의 트립체인을 분석하는데 이용할 수 있다.



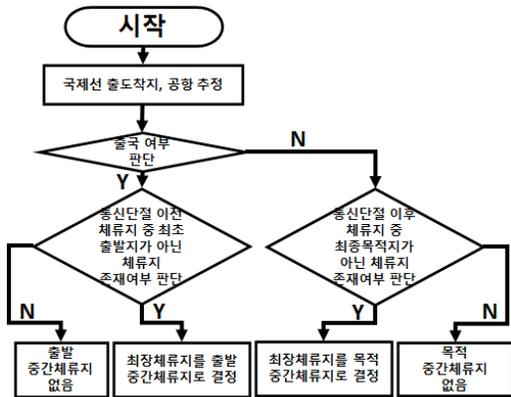
〈그림 12〉 중간체류지 추정 알고리즘

국제선 이용자의 경우에는, 국외에서의 이동 경로 추정이 불가하며 이에 따라 국내에서 발생한 통행만을 대상으로 중간체류지를 추정한다.

이를 위한 알고리즘은 국내선 이용자의 잠재지 및 중간체류지와 유사하게 정의한다.

〈표 2〉 국제선 중간체류지 추정 기준

대상	상세
출발중간 체류지	- 국내 출발, 해외 도착 경로 상의 국내 중간 체류지 - 출발지에서 출발공항까지 도착하기 전 일정시간 이상 체류한 지역 중 최장 체류지
목적중간 체류지	- 해외 출발, 국내 도착 경로 상의 국내 중간 체류지 - 도착공항에서 목적지까지 도착하기 전 일정시간 이상 체류한 지역 중 최장 체류지
체류시간 기준	- 20분 이상



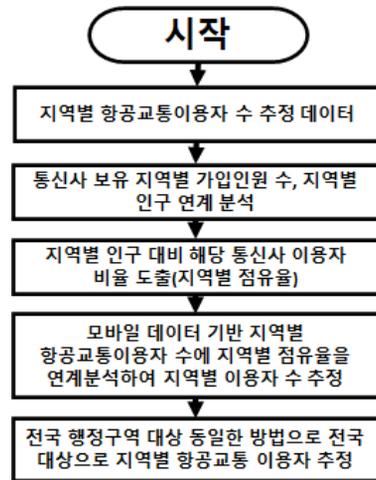
〈그림 13〉 국제선 중간체류지 추정 알고리즘

3.2.5 모수화

단일 통신사를 통하여 항공교통이용자 O/D 추정 시, 해당 통신사의 점유율로 인해 실제 숫자와 실제 이용자 숫자와 큰 차이를 보일 수 있다. 따라서 모바일 통신 기지국 데이터를 기반으로 한 항공교통이용자 데이터를 지역별 인구 및 점유율을 토대로 보정하는 모수화를 통해 국내 인구 전체를 대상으로 한 분석이 가능하도록 보완한다.

우선 각 통신사가 보유한 지역별 개통자 수 정보와 통계청에서 제공하는 지역별 인구를 토대로 지역별 점유율을 도출한다. 이렇게 추정된 지역별 점유율을 이용해 모수화를 위한 지역별 계수를 추출하고 해당 지역의 항공교통이용자에

계수를 적용하여 지역별 항공교통이용자를 추정한다. 이를 통해 실제에 가까운 항공교통이용자 수를 추정할 수 있다.



〈그림 14〉 모수화 알고리즘

3.3 데이터 추출 및 검증

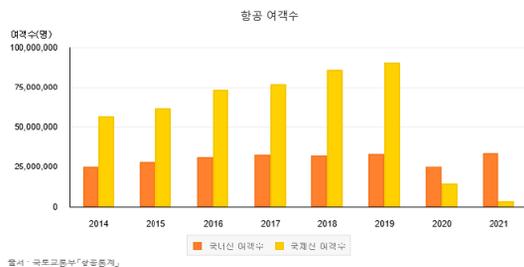
3.3.1 추출 범위

본 연구에서는 전국적인 항공교통이용자 및 이동경로 추정을 위하여 운영중인 모든 국내공항 및 전국 행정구역을 대상으로 O/D 데이터를 추출하였다. 추출대상 범위는 아래와 같다.

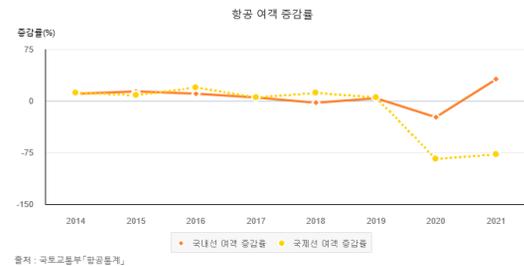
〈표 3〉 모바일 데이터 기반 항공교통이용자 O/D 추정 범위

범위	상세
시간적 범위	- 2019년 1~12월
공간적 범위	- (공항)전국 15개 공항(8개 국제공항, 7개 국내공항) - (지역)전국(행정동/시군구 수준에서 항공교통이용자 및 이동경로 추출)
대상 통신사	- 국내 주요 통신사 보유 데이터 * 30% 이상의 비교적 높은 점유율을 보유한 통신사 데이터 활용
국적 범위	- 내국인 및 외국인

시간적 추출 범위는 2019년 1~12월로, 2019년 말에 발생한 코로나19 팬데믹을 고려하여 결정하였다. 코로나19 대유행 이전에는 지속적으로 항공교통이용자수가 증가해 왔으나, 코로나19 대유행으로 인하여 2020년에는 국내선 23.7% 국제선 84.2% 가량 감소하였다.[9] 이에 따라 후속연구로 코로나19 대유행 종식 이후의 항공교통이용자 예측이나 코로나19 대유행에 따른 영향 분석에도 활용할 수 있도록 2019년을 대상으로 데이터를 추출하였다.



〈그림 15〉 연도별 항공여객수 추이(2014~2021)



〈그림 16〉 연도별 국제/국내선 여객증감률 추이 (2014~2021)

3.3.2 데이터 추출 및 검증

상기 알고리즘을 활용하여 전국의 15개 공항을 대상으로 항공교통이용자 O/D 데이터를 추출과 검증을 수행하였다. 우선 추출된 데이터는 행정구역 단위에 따라 시군구 단위, 행정동 단위로 각각 나누어 추출하였다. 추출된 데이터는 기준년, 내외국인, 연령대, 국제/국내선 구분, 출/도착 공항코드, 여객수 등의 데이터를 포함한다. 이때

공항은 IATA (International Air Transport Association, 국제항공운송협회)에서 규정한 공항코드로 표시하며, 각 지역은 행정안전부에서 제공하는 행정표준코드를 기준으로 표시한다.

A	B	C	D	E	F	G	H	I	J
ETL_Y	CNTRY_YN	AGE_CD	INT_AIRPORT_YN	ORIGIN	ORI_AIRPORT	DES_AIRPORT	destin	PSSGR_KT	PSSGR_CNT
2	2019	0	0	0	11560 GMP	PUS	48250 *	*	*
3	2019	0	0	0	48250 GMP	PUS	48250 *	*	*
4	2019	0	0	0	11560 GMP	KWJ	46810 *	*	*
5	2019	0	0	0	48250 PUS	GMP	11470 *	*	*
6	2019	0	0	0	11560 GMP	PUS	26500 *	*	*
7	2019	0	0	0	48250 PUS	GMP	11500 *	*	*
8	2019	0	0	0	11545 GMP	PUS	26200 *	*	*
9	2019	0	0	0	48250 PUS	GMP	11650 *	*	*
10	2019	0	0	0	11530 GMP	PUS	48127 *	*	*
11	2019	0	0	0	48310 PUS	CJU	NULL *	*	*
12	2019	0	0	0	11530 GMP	RSU	46130 *	*	*
13	2019	0	0	0	48310 PUS	GMP	11440 *	*	*
14	2019	0	0	0	NULL CJU	GMP	11440 *	*	*
15	2019	0	0	0	48310 PUS	GMP	41287 *	*	*
16	2019	0	0	0	NULL CJU	TAE	27230 *	*	*
17	2019	0	0	0	48310 USN	PUS	48310 *	*	*
18	2019	0	0	0	NULL CJU	CJU	36110 *	*	*
19	2019	0	0	0	48330 PUS	ICN	NULL *	*	*
20	2019	0	0	0	NULL CJU	GMP	41173 *	*	*
21	2019	0	0	0	48720 PUS	CJU	50110 *	*	*
22	2019	0	0	0	NULL CJU	KWJ	46790 *	*	*
23	2019	0	0	0	48730 PUS	GMP	11710 *	*	*

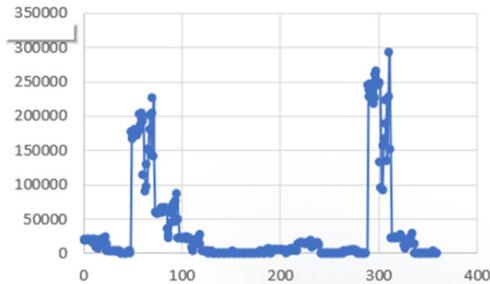
〈그림 17〉 항공교통이용자 O/D 추출 데이터 예시

검증의 경우, 본 연구에서는 상관성 분석을 통해 수행하였다. 설문조사를 통하여 항공교통이용자의 O/D를 직접적으로 조사하고 비교분석하는 것이 가장 확실한 방법이나, 소요되는 시간과 비용이 과대하며, 불성실, 무응답 등으로 정확도가 낮다는 한계가 있다. 또한 통신사 점유율 등의 문제로 인하여 모바일 통신데이터 기반 항공교통이용자 수가 실제와 큰 차이가 있다는 면에서 검증에 한계가 있다. 이에 따라 본 연구에서는 상관성 분석법을 통하여 모바일 통신 데이터 기반 항공교통이용자 O/D 데이터를 검증하였다.

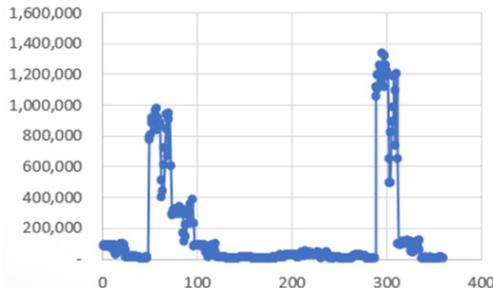
상관성 분석법은 확률/통계에서 두 변수 간에 어떤 선형적 관계를 가지고 있는지를 분석하는 방법으로 양쪽의 변화패턴의 일치성을 비교분석하여 항공교통이용자 O/D 데이터의 정확성을 검증한다. 이때 상관관계(R^2)가 1에 가까우며 추세가 유사할 경우 적정데이터라 판단하게 된다. 본 상관성 검토의 비교데이터는 항공통계 기반의 월별 항공교통이용자 수 데이터를 활용하였다.

검증은 월별, 공항별로 추출된 모수화되지 않은 항공교통이용자 O/D 데이터와 항공통계 기반 항공교통이용자 데이터의 그래프 기울기를 비교

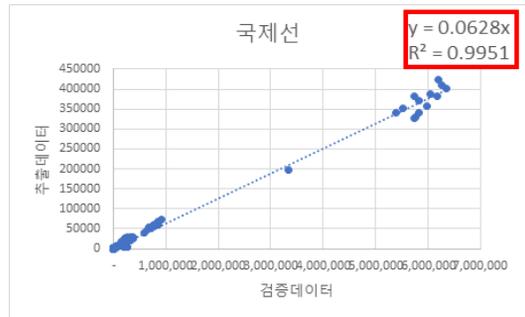
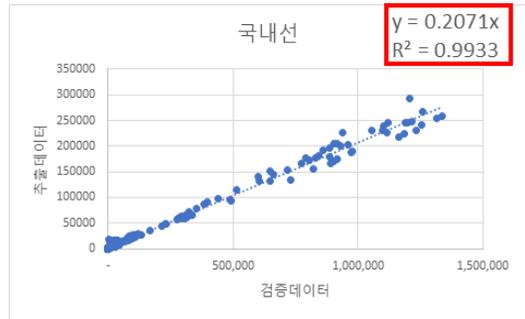
분석하는 방식으로 수행되었다. 분석한 결과, 모바일 통신데이터 기반 항공교통이용자 추정치와 항공통계 기반 항공교통이용자 통계값 양쪽의 패턴이 유사한 것으로 확인되었으며, 상관관계 역시 높은 것으로 분석되었다. 국내선의 경우, 결과값이 다소 차이를 보이거나(0.2) 상관관계가 높은 것으로 분석되었고, 국제선의 경우, 결과값의 차이가 적고(0.06), 상관관계가 높은 것으로 분석되었다. 전반적으로는 상관성이 0.99로 도출되었으며, 통신데이터 기반 항공교통이용자 추출 데이터를 적정데이터로 판단하였다.



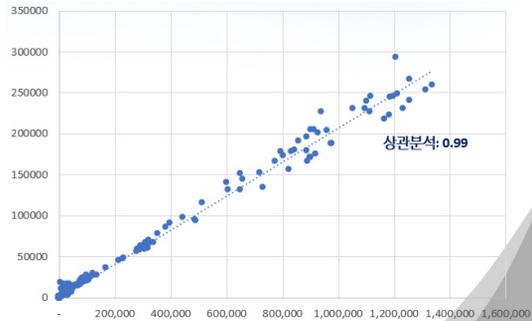
〈그림 18〉 통신데이터 기반 월별 항공교통이용자 패턴
(x축 : 시간(일 단위), y축 : 이용자 수(명))



〈그림 19〉 항공통계 기반 월별 항공교통이용자 패턴
(x축 : 시간(일 단위), y축 : 이용자 수(명))



〈그림 20〉 국제선/국내선 상관관계 분석



〈그림 21〉 통신데이터-항공통계 상관관계 분석

IV. 결론

본 연구에서는 기존의 통계적인 방법을 통한 항공교통이용자 분석의 한계를 넘어 각 공항별 이용자의 출/도착지역 범위, 이용자별 이동경로 등 항공교통이용자의 통행형태 전반을 분석하기 위한 기초데이터 생성을 위해 모바일 통신 데이터를 기반으로 항공교통이용자 O/D를 추출하였다.

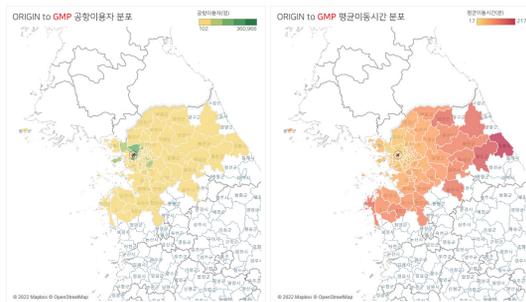
모바일 통신 데이터 기반의 항공교통이용자 O/D 데이터는 기존의 통계자료와 다르게 각 항공교통이용자의 이동경로, 출발지 등 이동행태 전체를 분석할 수 있는 기초자료가 될 수 있다. 특히 각 공항에 대한 공항접근영향권(Catchment Area, 각 공항이 가지는 지리적 영향범위)을 분석할 수 있으며 이를 통해 각 공항별 수요예측, 주변지역 수요 검토 등에 활용할 수 있다. 뿐만 아니라 지역별 공항이용자 수, 신공항 후보지 검토, 지역별 국제선 해외 목적지 분석, 출발지 공항선택 행태 분석 등에 활용할 수 있다. 따라서 향후 모바일 통신데이터 기반 항공교통이용자 O/D 데이터의 다양한 활용방안에 대한 후속연구

가 필요하다.

또한 보정계수를 활용하여 보다 정확한 모수화를 수행하여 실제 인구에 가까운 분석을 수행할 수 있도록 데이터를 보정하는 방법도 연구할 필요가 있다. 보정계수는 항공교통이용자 O/D 데이터를 모수화, 유동인구 보완, 외국인 수요 등을 전인구 규모로 보정하여 다양한 분석을 수행하기 위한 수단으로, 본 연구 과정에서 만든 약 5종의 보정계수를 간략히 설명하면, 우선 보정계수 1은 모수화 계수이다. 모수화는 상기 모수화 알고리즘과 같이 통신사업자의 지역별 점유율, 지역별 인구를 고려하여 각 지역마다 다르게 적용되며 본 연구의 보정과정에서도 동일하게 적용되었다. 그러나 해당 지역별 점유율은 통신사업자의 대외비에 해당하는 정보로 상세히 기술하기 어려워하기의 보정계수에는 모수화 보정계수 평균치를 기술하였다. 두 번째로 보정계수 2는 유동인구 기준 보정으로, 통신사의 점유율을 반영하여 전체적인 통행수를 보정하는 계수이다. 보정계수 3은 김포/인천 등 주요 공항의 모바일 통신 기반 항공교통이용자 수와 항공통계기반 실측치와 비교하여 만든 보정계수이다. 보정계수 4는 내외국인 구분을 위한 것으로 법무부의 출입국자 통계 데이터와 모바일 통신데이터 기반 국제선 이용자 수를 비교하여 생성한 보정계

〈표 4〉 모바일 데이터 기반 항공교통이용자 O/D 활용 방안 예시

범위	상세
공항접근영향권 분석	- 각 공항별 기능과 역할 검토 - 공항별 수요예측시 주변지역의 수요 예측
지역별 공항이용자 수	- 공항주변지역 여객분석, 공항 접근시 공항별 교통수단 확충 및 운영개선
신공항 후보지 검토 및 잠재 수요 분석	- 신공항 개발 검토시 해당 지역 여객 수요 파악, 잠재 수요 검토
지역별 해외 목적지 분석	- 각 지역별 국제선 여객이 도착지(국가, 도시) 분석 - 국제선 수요 분석 및 노선개발 등 활용



〈그림 22〉 2019년 김포공항 이용자 분포 및 평균이동시간

〈표 5〉 보정계수 예시

구분	기반 데이터	국내선 적용	국제선 적용
보정계수1	모수화 비율	5	16.5
보정계수2	유동인구 기반	3.5	3.5
보정계수3	공항실적, 통신데이터	2	12.5
보정계수4	법무부 입출국 데이터, 통신데이터	-	(내국인) 10.94 (외국인) 1.26
보정계수5	보정계수3(12.5)+ 보정계수4(1.26)	-	13.76

수이다. 보정계수 5는 국제선을 이용한 외국인 수를 보정하기 위한 것으로, 보정계수 3의 국제선 수치와 보정계수 4의 외국인 보정계수를 더하여 생성한 보정계수이다.

다만, 해당 보정계수를 적용하여 보정한 모바일통신데이터 기반 항공교통이용자 수가 항공통계값과 많은 차이를 보였으며, 해당 계수는 많은 개선이 필요할 것으로 판단된다. 따라서 보정계수를 통한 보정방안의 개선과 보정계수의 보완 역시 향후의 과제로 한다.

참 고 문 헌

- [1] 항공통계매뉴얼, 2013.
- [2] 2021년 항공운송사업전망, 한국교통연구원, pp. 9, 2021.
- [3] Yuanyuan Qiao et al., "A Mobility Analytical Framework for Big Mobile Data in Densely Populated Area", IEEE Transactions on Vehicular Technology, 66(2), pp. 1443-1455, 2017.2.
- [4] 유정훈, 집계 관측 자료를 이용한 국제선 여객 공항선택 모형, 대한교통학회지 36(5), pp. 392-401, 2018.10.
- [5] 박재성 외, 시계열 자료의 안정성을 고려한 항공교통 계량경제모형 개발, 대한교통학회지 34(1), pp. 95-106, 2016.1.
- [6] 김제철, 손홍구, 박진서, ARIMA-Intervention 모형을 통한 국제선 항공수요예측 연구, 대한교통학회지 37(1), pp. 51-65, 2019.2.
- [7] 김영록, 한국의 국제선 항공수요 예측과 검토, 한국항공운항학회 27(3), pp. 98-105, 2019.1.
- [8] 미래교통수요의 변화예측과 대응, 한국교통연구원, 2017.11.30.
- [9] 연도별 항공여객 수송 실적, 통계청, 2023.
- [10] 데이터로 보는 한국관광 2020, 한국관광공사, pp. 20, 2020.

저 자 소 개



조 범 철(Bumchul Cho)

· 2009년 2월: 서울대학교 환경 계획학과 (박사)
 <관심분야> 모빌리티 분석, 모빌리티 빅데이터, 교통 빅데이터 플랫폼



권 기 훈(Kihun Kwon)

· 2008년 8월: 한국항공대학교 정보통신공학과 (공학사)
 · 2014년 2월: 한국항공대학교 항공전자공학과 (공학석사)
 <관심분야> 모바일 통신 빅데이터 활용, 도시로봇 등