

활동기반 시뮬레이터 입력 자료의 전처리 방안에 대한 연구: FEATHERS Seoul을 사례로

조성진¹ · 황정환² · Tom Bellemans¹ · Bruno Kochan¹ · 이원도² · 최기주³ · 조창현^{2*}

¹ Hasselt University IMOB, ² 경희대학교 지리학과, ³ 아주대학교 교통시스템공학과

A Study on Data Preprocessing for the Activity-Travel Simulator: A Case of FEATHERS Seoul

CHO, Sungjin¹ · HWANG, Jeong Hwan² · BELLEMANS, Tom¹ · KOCHAN, Bruno¹ ·
LEE, Won Do² · CHOI, Keechoo³ · JOH, Chang-Hyeon^{2*}

¹ IMOB, Hasselt University, 3590 Diepenbeek, Belgium

² Department of Geography, Kyung Hee University, Seoul 130-701, Korea

³ Department of Transportation Systems Engineering, Ajou University, Gyeonggi 443-749, Korea

Abstract

Research on activity-based travel demand forecasting and activity-travel simulator has received an international attention for the last two decades. Ways to develop the activity-based simulator may be manifold. It is obvious that importing an existing simulator that has been proven internationally likely reduces the development cost and the risk of failure. By definition of the activity-based approach, however, the details of an activity-based simulator inevitably relies on particular social, economic and cultural characteristics of the society where the simulator is developed. When importing such a simulator from overseas, the researcher should be aware of the importance of tuning the system for the society to which the imported system is applied. There are many potential works on this, including for example the tuning of data structure that is likely different form of the original system. The authors are yet aware of certain research on those. The current paper aims to report the result of transforming the input data for applying the existing activity-travel simulator to Seoul. The paper first introduces FEATHERS that was developed in Belgium having Albatross which is the core of system. FEATHERS Seoul that is under development and modified version of the original FEATHERS is briefly described and the related problems are discussed. The paper then explored to resolve and to alleviate such problems.

교통수요 예측을 위한 활동기반 이론과 그 실행을 위한 시뮬레이터에 대한 연구가 국내외적으로 활발하다. 활동기반 시뮬레이터의 개발 방법 중 다른 사회·경제·문화적 토대로부터 기 개발된 시뮬레이터를 새로운 연구지역에 전용하는 방법(model transferability)은 개발 비용이 저렴하고, 실패 확률이 적다는 장점이 있다. 하지만 활동기반 교통수요 이론의 정의로부터, 시뮬레이터 구축 자체는 연구지역의 사회·경제·문화적 특성에 특화되어 있으므로 먼저 시뮬레이터의 자료구조와 구성요소에 대한 맥락적 이해가 선행되어야 한다. 이 같은 중요성에도 불구하고 활동기반 시뮬레이터의 공간적 전용 시 발생하는 문제들에 대한 관련 연구는 아직까지 보고된 바 없다. 이 논문은 ALBATROSS 스케줄 엔진을 핵심으로 하는 FEATHERS 활동-통행 시뮬레이터를 한국의 수도권에 적용하기 위해 필요한 입력 자료의 전처리 과정에 대한 연구 결과를 자료구조와 구성요소를 중심으로 보고한다. 이를 위해, FEATHERS 시스템과 입력 자료를 간략히 소개하고, FEATHERS Seoul 플랫폼의 개발 시 발생하는 문제를 조사하고, 그 해결책과 적용 결과를 제시한다.

Keywords

activity-based approach, activity-travel simulator, FEATHERS Seoul, input data, Seoul metropolitan area, travel demand
활동기반 접근법, 활동-통행 시뮬레이터, FEATHERS Seoul, 입력 자료, 수도권, 교통수요

* : Corresponding Author
bwchjoh@khu.ac.kr, Phone: +82-2-961-9264, Fax: +82-2-961-0848

Received 23 July 2014, Accepted 14 October 2014

서론

교통수요 예측을 위한 활동기반 이론과 그 실행을 위한 시뮬레이터에 대한 연구가 국내외적으로 활발하다 (Pinjari and Bhat, 2010; Lim et al., 2013). 활동기반 이론은 통행이 활동으로부터 파생되기에 개인과 가구의 일상 활동 운영 원리를 그들 사회와의 관계를 통해 이해하는 것이 선행된다는 도시계획학자 Chapin과 시간지리학자 Hägerstrand의 도시 서비스에 대한 수요와 공급에 관한 이론에서 기원한다(Timmermans et al., 2002). 활동기반 이론으로부터 개발되는 활동-통행 시뮬레이터는 일반적으로 활동-통행의 의사결정의 환경이 되는 장단기 조건, 인구 전수화, 활동-통행 시뮬레이션, 예측 평가, 결과 보고 등의 모듈들로 구성된다(Arentze and Timmermans, 2012). 이러한 모듈들을 구성하는 이론들은 매우 많은데, 특히 의사결정 이론, 학습과 업데이트 및 불확실성 관련 이론, 자동화 자료수집 방법, 활동 추론 방법, 활동-통행의 공간분석 방법, 토지이용과 교통 간 관계 관련 이론, 도시 형태와 통행 형태 관련 이론 등이 중요한 토픽으로 연구되고 있다.

최근 이러한 활동-통행 시뮬레이터 중 가장 완성도가 높은 것들은 실제 특정 나라의 중앙이나 지방 정부 교통 정책에 활용되기도 하는데, 독창성이 높은 시뮬레이터의 많은 사례 중 CEMDAP(Bhat et al., 2004)과 ALBATROSS(Arentze et al., 2000)의 활동-통행 의사결정 엔진을 각각 채택한 SimAGENT(Goulias et al., 2012)와 FEATHERS(Bellemans et al., 2010)가 주목된다. CEMDAP-SimAGENT는 계량경제학적 효용극대화 의사결정 원리에 기초한 활동-통행 시뮬레이터이며, ALBATROSS-FEATHERS는 인지심리학적 규칙기반 의사결정 원리에 기초한 활동-통행 시뮬레이터이다. 전자는 미국 서부 캘리포니아의 SCAG에서 활동기반 교통정책 실행을 법제화한 것에 근거하여 개발한 시뮬레이터이다. 후자는 벨기에 북부 플란더스 지역의 다양한 교통수요 관리 정책의 영향을 검토하기 위해 개발한 시뮬레이터이다. 각 시뮬레이터의 장단점은 각각 기초하고 있는 계량경제학적 의사결정 이론과 인지심리학적 의사결정 이론으로부터 확인될 수 있으며, 이에 대한 논의는 현재 진행형이다.

본 연구는 비교적 구조가 간결하고, 모델의 전용성이 뛰어나며, 자료의 요구조건이 유연한 FEATHERS를 시뮬레이터로 채택하였다. FEATHER가 핵심 엔진으로

채택하고 있는 ALBATROSS 규칙기반 모형의 의사결정나무 데이터마이닝 기법은 시뮬레이션 결과의 해석이 용이하다. 또한 FEATHERS의 모듈화된 구조는 전혀 새로운 지역에 적용할 때 필요한 모듈만 개별적으로 수정할 수 있어 전용성이 우수하다. 특히 플란더스 등 구체적인 지역의 다양한 정책 평가에 필요한 평가 모듈들이 이미 탑재되어 있어 활용가능성이 높다.

우리나라에서도 교통수요 관리 측면의 교통 정책이 강조되는 현재(Lee et al., 2011; Park et al., 2012), 활동기반 활동-통행 시뮬레이터의 개발과 이를 이용한 교통수요 관리 정책 효과의 시나리오 분석은 그 학문적 성과 외에도 실효성 측면에서 기대되는 바가 크다. 실제로 FEATHERS 플랫폼을 예로 들면, 플란더스 지방의 재택근무 시나리오에 따른 승용차통행 감소 예측(Kochan et al., 2011), 교통사고 위험 노출의 교통준별 추정(Pirdavani et al., 2012), 전기자동차 충전의 시계열적 수요 예측(Knappen et al., 2012) 등을 들 수 있다. 따라서 본 연구팀은 이러한 학술적 및 실무적 필요성 충족에 기여하기 위하여, FEATHERS 플랫폼의 원개발자인 벨기에 하셀트 대학 교통연구소(IMOB)와의 국제 공동 연구 형태로 FEATHERS 플랫폼의 한국형 시스템 FEATHERS Seoul(이후 FS)의 개발 연구 중에 있다. 이는 이 연구에서 FEATHERS를 채택한 또 하나의 이유가 되겠다.

활동기반 시뮬레이터의 개발 방식으로는 크게 두가지로, 시스템 설계, 개발, 보완의 전 단계에 걸친 자체 개발 방법과, 기존에 개발된 시뮬레이터를 전용하는 방법을 들 수 있다. 후자는 사회·경제·문화적 토대가 다른 곳으로부터 기 개발된 시뮬레이터를 연구지역의 실정에 맞추어 수정하는 방법으로, 전자에 비해 개발 비용이 저렴하고, 개발에 필요한 시간을 단축할 수 있으며, 실패 확률이 작다. 하지만 활동기반 교통수요 이론의 정의로부터, 시뮬레이터의 구축 과정은 연구지역의 사회·경제·문화적 특성에 특화되어 있기 때문에 새로운 연구지역에 대한 전용 시 시스템의 자료구조 및 구성요소에 대한 이해가 선행되어야 한다. 이러한 중요성에도 불구하고 교통수요 예측을 위한 활동기반 시뮬레이터의 전용 시 발생하는 문제들에 대한 실증 연구는 아직까지 보고된 바 없다(Arentze et al., 2002; Siker et al., 2013). 따라서 이 논문은 FEATHERS 활동기반 교통수요 예측 시뮬레이터를 한국의 수도권(SMA: Seoul Metropolitan Area)에 적용하기 위한 FS 개발에 있어서 자료구조의

불일치, 구성요소의 해석상의 차이로 인해 발생하는 문제점과 그 해결방안으로서 자료 전처리 방법을 제안하고, 그 적용결과를 실증연구를 통해 보고한다.

이 논문의 구성은 다음과 같다. 2장에서는 FEATHERS 시스템과 입력 자료에 대해서 간략히 소개하고, 다음 3장에서는 FS 플랫폼 개발 시 발생하는 문제점에 대해서 입력 자료의 구조와 구성요소의 차원에서 논의한다. 곧이어 4장에서는 각 문제점에 대한 대안과 그 적용결과를 제시한다. 마지막 5장에서는 연구 결과를 요약하고 향후 과제를 제안함으로써 본 논문을 마무리한다.

FEATHERS 시스템과 입력자료

본 연구에서 적용한 FEATHERS는 교통수요 예측을 위한 시뮬레이션 시스템으로, 시스템의 활용성 및 전용성을 확보하기 위해 핵심 모듈인 ALBATROSS 스케줄 엔진을 중심으로 각각의 기능을 가진 여러 모듈들의 집합체로 구성되어 있다(Bellemans et al., 2010). 세부적으로 Configuration 모듈은 시스템 관리 전체를 관장하는 콘트롤 타워 역할을 하며, XML 파일을 이용하여 새 모듈의 추가 시 새 파라미터 세팅을 추가하는 등의 방법으로 이용자가 손쉽게 수정, 관리할 수 있도록 되어 있다. DatMod 모듈은 시뮬레이션에 필요한 교통 공급(존, 네트워크, 입지 자료 등) 및 수요(활동-통행) 자료를 관리한다. PopMod 모듈은 인구센서스 상의 인구 특성 자료를 관리하며, 기준년도 인구정보를 관리하고, 목표연도의 전수화인구를 생산한다. ScheMod은 관측 자료를 토대로 활동-통행 의사결정규칙을 정의한다. 현재 FEATHERS는 ALBATROSS의 의사결정나무(DT: Decision tree)를 이용한 규칙기반모형을 채택하고 있다. ExecMod는 ScheMod에서 정의된 의사결정규칙과 교통시스템의 관계를 정의하며, 또한 스케줄 실행에 따른 결과를 평가하여 기존 스케줄에 피드백될 수 있도록 구현한다. LearnMod은 반복적인 스케줄 혹은 장기적인 교통 여건 자체의 변화 등에 의해 의사결정규칙의 변화를 가능하게 한다. StatMod와 VisMod은 활동-통행 시뮬레이터의 결과를 간단한 통계수치와 지도로 표현해 모의실험 결과를 직관적으로 이해하는 데 도움을 준다. 마지막으로 TrainMod는 시나리오 기반 시뮬레이션이 가능하도록 시스템을 실제 자료에 적용하는 기능을 한다.

이상 각각의 모듈들은 상호간 유기적으로 연계되어

있으며, FEATHERS 플랫폼으로서 통합적인 기능을 제공하고 있다. 이 같이 모듈화된 시스템은 전체 시스템구조의 변경 없이 사용자의 필요에 따라 각 모듈들을 개별적으로 수정 및 대체할 수 있기 때문에 시스템의 전용이 용이하다. 그럼에도 불구하고 모델 내부에서 정의하고 있는 여러 활동-통행 요소들은 원개발지역인 유럽 실정에 맞춰져 있기 때문에 이를 우리나라의 실정에 맞도록 모델을 수정하는 노력이 요구된다.

FEATHERS의 입력 자료는 크게 개인의 활동-통행 다이어리, 마이크로 인구 자료, 환경 자료로 구성된다. 세부적으로, 다이어리 자료는 가구 및 가구원(개인)의 사회·인구학적 특성과 더불어, 각 개인이 일과시간 동안 수행한 활동-통행 정보 등을 나타낸다. 기존 FEATHERS는 벨기에에서 4년 마다 실시하는 가구통행실태조사(이후 가통: HTS)의 일종인 OVG(Onderzoek Verplaatsings Gedrag)에서 다이어리 자료를 획득하여 입력 자료로 활용하고 있다. OVG는 플란더스 지방에 거주하는 8800명 개개인의 활동-통행 다이어리 정보를 포함하고 있다.

다음으로 마이크로 인구 자료는 연구지역의 전수 인구에 해당하는 가구, 가구원 정보를 포함하고 있으며, 앞서 다이어리 자료의 구조와 동일하다. 일반적으로 마이크로 인구 자료는 직접 수집이 어렵거나, 사용상 제약이 많기 때문에 일반적으로 인구전수화기법을 이용해 자료를 직접 구축한다. 이러한 이유로 FEATHERS는 OVG자료와 플란더스 지방의 인구센서스 자료를 IPF(Iterative Proportional Fitting) 알고리즘에 적용하여 목표연도의 인구 자료를 생산하여 적용하고 있다(Guo and Bhat, 2007).

마지막으로 환경 자료는 Superzone, Zone, Subzone의 계층적 공간체계에 따라 토지이용과 네트워크 정보로 구성되어 있다. 먼저 토지이용 정보는 Superzone을 제외한 각 공간단위별로 다시 구분하는데, Zone 단위에서는 각 구역별로 주차시설과 관련된 정보(장기/단기 주차 시설 수, 장기/단기별 주차비용 등)를 포함하고 있으며, Subzone 단위는 각 구역 도시 밀도, 면적, 총 가구 수 등의 일반 현황과 함께 각 산업 영역별 종사자 수에 관한 정보를 포함하고 있다. 이 산업 영역별 종사자 정보는 FEATHERS의 각 활동 유형에 따른 활동 공간 선택에 있어서 공간단위별 흡입력으로서 작용한다. 다음 네트워크 정보는 출발지와 도착지의 OD 행렬의 자료구조를 가지고 있으며, 각 OD에 해당하는 통행 수단별 통행 소요 시간, 통행 거리, 비용 등에 관한 통행 서비스 정보를 포함하고 있다.

FEATHERS Seoul의 입력자료 문제

수도권의 통행수요 분석 및 예측을 위한 FS 플랫폼을 구현하기 위해서는 먼저 FEATHERS의 입력자료 요구 조건을 충족시켜야 한다. 따라서 본 장에서는 연구지역에서 이용 가능한 자료를 살펴보고, 입력 자료의 전처리 과정에서 발생하는 여러 문제점에 대해서 자료구조 및 구성요소의 측면에서 논의한다.

먼저 2010년 수도권에서 시행된 가통 자료로 부터 (수도권 전체 인구의 2.5%에 해당하는) 가구 및 가구원의 사회·인구학적 특성과 활동-통행 정보를 추출하여 FS의 활동-통행 다이어리를 구축하였다. 이 가통 자료는 가구, 가구원 항목과 조사기간 동안 응답자에 의해 수행된 통행 관련 항목으로 구성되어 있는데, FS의 다이어리와 전반적으로 유사한 자료구조를 가지고 있지만(조사 명칭에서도 알 수 있듯이) 조사 목적상 통행 정보를 중심으로 수집하기 때문에 이 자료에서 활동 정보를 직접 추출할 수 없다. 더군다나 가통 자료의 일부 항목에서는 FS의 항목과 차이가 나타난다. 위 사례를 다음의 대표적인 유형으로 분류하였다.

- 유사 항목임에도 자료의 형태가 다른 경우
- 논리적으로 자료의 직접 적용이 불가능한 경우
- 필요 항목이 없는 경우

첫째, 가통 자료에서 FS 입력 자료와 유사한 항목이 존재하지만 자료 형태가 다른 경우의 예시로, 가통 조사에서는 응답자의 연령을 출생년도의 연속형 자료로 표기하는 반면에 FS는 미리 설정된 급간에 따라 연령대의 범주형 자료로 정의한다. 이 경우는 비교적 단순한 문제로 가통 자료를 FS의 자료형으로 변환시키기만 하면 된다. 단순한 문제임에도 자료형의 불일치는 시스템 실행상 문제를 일으킬 소지가 있으므로 반드시 사전에 처리되어야 한다.

둘째, 논리적 근거로 인해서 가통 자료를 직접 활용할 수 없는 대표적인 사례로, FS는 24시간 동안에 거주지를 기반으로 하는 발생하는 일일 활동-통행에 대한 정보를 사용하지만, 가통 자료는 조사과정에서 이러한 엄격한 시공간적 제약조건이 존재하지 않기 때문에 24시간을 초과하거나 최초 출발지, 최종 목적지가 거주지가 아닌 자료들이 다수 포함되어 있다. 이 같은 FS 입력 자료의 요건에 위배되는 자료들은 시뮬레이션 실행시 오류를

발생시킬 수 있다. 따라서 사전에 자료 요건을 설정하고 이를 전처리 과정에 적용해 오류를 미연에 방지해야 한다.

셋째는 가통에서 FS의 입력자료 항목을 추출할 수 없는 경우로, 통행 실태의 파악을 위해 시행된 가통 자료에서 활동에 대한 정보를 직접 획득할 수 없는 것이 이에 해당된다.

이상의 FS 다이어리 자료의 획득 이외에도 중요한 전처리 과정이 하나 더 있는데, 이는 연구지역의 전수인구 자료를 준비하는 과정이다. 일반적으로 이 전수인구 자료는 획득이 어렵거나 불가능한 경우가 많아 대신에 IPF(Iterative Proportional Fitting)나 CO(Combinatorial Optimization)등과 같은 인구전수화 방법을 이용해서 참조자료와 마진자료를 토대로(가능한 실제에 가까운) 전수인구 자료를 생산한다. 이에 FS는 2010년 가통자료의 가구 및 개인(가구원)의 인구정보와 수도권교통본부에서 구축한 2010년 사회경제지표를 각각 참조자료와 마진자료로 이용하였다. 하지만 참조자료의 가구와 가구원의 각 항목별 빈도를 순차·반복적으로 마진자료의 각 단위의 항목별 총계에 맞추어 나가는 IPF의 특성상, 마진자료로 이용한 사회경제지표에는 FEATHERS의 가구단위 항목에 해당하는 정보가 없기 때문에 전수인구 자료 중 가구단위 정보의 전수 정확도를 보장할 수 없다. 이 문제를 해결하기 위해서 본 연구에서는 참조자료(가통 자료)의 가구단위 항목과 일치하고 국내에서 획득 가능한 주택점유형태(자가, 전세, 월세 등)와 주택유형(아파트, 단독주택, 빌라, 오피스텔 등) 항목을 가구단위 마진정보로 이용했다. 또한 IPF와 비교해 리스트 기반의 자료구조로 전처리 과정이 단순하고, 실행속도가 빠르며, 상대적으로 전수 정확도가 높은 IPU (Iterative Proportional Updating)를 이용하여 FS의 전수인구 자료를 생산하였다. 특히 플란더스의 OVG 자료의 경우와 달리, 수도권의 가통 자료는 센서스와 일치하는 것이 항목이 별로 없어 이 같은 자료 문제에도 다소 높은 정확도를 보장하는 IPU를 사용하였다. IPU에 대한 보다 상세한 내용은 다음 장에서 소개하겠다.

연구지역에서 FEATHERS의 토지이용 자료 중 Zone 단위에 대응하는 주차시설 정보를 획득할 수 없기 때문에 FS는 수도권의 79개 시군구 단위를 Superzone에, 1107개의 행정동 단위를 Zone과 Subzone에 각각 적용하여 실질적으로 2단계의 공간단위체계를 채택하였다(Lee et al., 2014). 참고로, 환경 자료 중 토지이용 정보는 수도권교통본부에서, 통행 서비스 정보는 4단계

교통수요 예측모형을 이용해서 정산한 네트워크 자료에서 각각 획득하였다.

위 경우에서 보듯이 가통 자료를 FS의 입력 자료로 이용하기 위해서는 사전에 입력 자료 전처리 과정에서 발생할 수 있는 문제들에 대해 예상하고, 이에 대한 명확한 처리 기준이 필요하다. 다음 장에서는 각각의 입력 자료 문제에 상응하는 자료 처리 기준 및 자료 불일치 완화 방안 등을 제안하고, 적용 결과를 실증 자료를 통해서 검증한다.

FEATHERS Seoul의 자료 전처리 과정과 그 효과

FS의 입력 자료를 생성하기 위한 가통 자료의 전처리 방법은 다음과 같다. 첫째, 가통 자료의 통행 요소를 FS의 활동 입력 자료로 직접 변환하는 방법으로, 통행 목적을 활동 유형으로 통행 시간 및 위치를 활동 시간 및 위치로 직접 변환한다. 둘째, 두 자료 간 항목의 정의나 분류 체계가 다른 경우에 목표 자료(여기서 FEATHERS 입력자료)의 형식에 최대한 일치시킨다. 구체적으로 가통 자료는 배움, 귀가, 출근, 등교, 학원, 업무, 직장복귀, 쇼핑, 여가·오락·외식·친지방문, 기타 활동의 총 10개 활동 항목을 구분하는데 반해서, FEATHERS는 쇼핑 활동을 일상과 비일상 쇼핑으로 구분하고, 여가, 친지방문, 관광을 각각 개별 활동 유형으로 구분한다. 통행수단의 경우, 가통 자료는 도보, 승용차 운전, 승용차 동승, 여러 대중교통 수단들 등 모두 18개의 항목을 구분하는데 반해서, FEATHERS는 승용차 운전, 승용차 동승, 대중교통, 도보의 4개 항목으로 구분한다. 마지막은 연관된 항목에서 필요한 정보를 유추하는 방식으로, 예를 들어 연속된 통행시간 정보에서 다음 활동시간을 유추한다.

1. 가구 및 가구원 자료

가통 자료로부터 추출되는 가구와 가구원 관련 세부 항목은 다음과 같다.

- 가구 구성(가구원 수, 직장인 수)
- 가구 월별소득
- 가구 내 최고령자/최연소 가구원의 연령대
- 총 보유차량 수
- 가구원 연령
- 직업 유무

- 성별
- 자동차 운전면허 보유 여부

전처리 과정은 가통 자료를 FEATHERS의 급간에 맞추어 변형하거나, 관련 정보를 합산하여 필요한 정보를 얻거나, 또는 이런 별도의 변환 과정 없이 그대로 적용한다. 다음은 각 처리 방식의 예시이다.

- i. 급간 적용의 예: 가통의 가구 내 최고령자 연령은 출생년도로 표기되는 반면에, FEATHERS는 18-34세, 35-54세, 55-64세, 65-74세, 75세 이상의 5개 급간을 설정하므로 가통의 출생년도 값을 2010년(조사년도)을 기준으로 연령을 계산하고 이를 FEATHERS의 급간에 맞추어 연령 값을 설정한다.
- ii. 정보 합산의 예: 가통은 승용차, 승합차, 소형화물차, 대형화물차 등 다양한 차종에 걸쳐 차량 보유 대수를 조사하고 있으나, FEATHERS는 0대, 1대 보유, 2대 이상 보유로 한정한다. 따라서 가통의 차종과 무관하게 보유 대수를 기준으로 가구 내 보유 차량수를 설정한다.
- iii. 직접 적용의 예: 가통의 성별, 승용차 운전면허 보유 여부 등 별도의 전처리과정 없이 FEATHERS에 직접 적용한다.

FEATHERS는 주어진 급간 별로 활동-통행 의사결정 규칙을 정의하기 때문에 각 자료의 급간을 설정하는 과정에 주의가 요구된다(Arentze et al., 2000). 하지만 실험결과 FS의 급간 변경에 따른 시뮬레이션 결과의 차이는 미미한 것으로 나타났다. 그럼에도 불구하고 새로운 연구지역에서 각 자료의 활동-통행 관련 의사결정 규칙에 대한 영향 정도를 사전에 평가하여 적합한 급간을 설정하는 작업은 필요하다.

한편 FEATHERS는 성인 남녀(가구주와 배우자)만을 분석 대상으로 설정하므로 가통 자료에서 대상 외의 정보는 직접 이용되지 않고, 다만 가구 내 최고령자 또는 최연소자의 연령 정보와 같은 가구 정보를 유추하는데 간접적으로 활용된다.

2. 활동 및 통행 자료

앞서 언급했듯이, FEATHERS는 개인의 통행 정보

(통행 시간, 통행 소요 시간, 통행 수단, 출발지, 도착지) 뿐만 아니라 활동 정보를 입력 자료로 사용하고 있다. 하지만 가통조사는 조사 목적상 활동 정보를 명시적으로 포함하고 있지 않기 때문에 통행 자료로부터 다음의 활동 정보를 유추해야 한다.

- 활동 유형
- 활동 시작 시간(시, 분)
- 활동 지속 시간(분)
- 활동 장소

활동시간을 예로 들면, 일련의 연속된 두 통행의 사이를 활동으로 간주하고, 두 통행의 시간 정보에서 활동시간을 유추할 수 있다. 따라서 이전 통행의 목적지는 이어진 활동의 유형으로, (이전)통행의 목적지는 활동 장소로 적용한다. 가통의 통행 목적과 FEATHERS의 활동유형 사이에 일부 불일치 항목에 대해서는 가통에서의 자료 가용 여부와 연구지역에서의 활동유형별 중요도를 참고하여 항목 간 병합 및 변형하여 FS에 적용한다. 그 처리 결과는 다음과 같다.

- # {통행 목적} → {활동 유형}: “{처리 방법}”
- 귀가 → 재택: “유지”
 - 출근 → 근무: “유지”
 - 외근 → 업무: “유지”
 - 배웅/마중 → 자녀등하교(bring/get): “유지”
 - 쇼핑 → (정기) 쇼핑: “병합>(*비정기 쇼핑 제외)
 - 학원 → 서비스: “변형”
 - 친목/여가/관광 → 여가활동: “병합”
(*친목, 관광 제외)
 - 기타 → 기타: “유지”

위 활동유형 자료의 처리과정에서 쇼핑 활동의 경우, FEATHERS에서는 활동의 주기성에 따라 정기·비정기로 구분하고 있지만, 가통자료에서 두 활동을 구분할 수 없기 때문에 가통에서 나타난 쇼핑 통행의 특성상 모두 정기 쇼핑으로 간주하였고, 따라서 비정기 쇼핑은 분석대상에서 제외시켰다. 마찬가지로 가통에서 구분할 수 없는 친목·여가·관광의 통행목적을 연구지역에서 상대적으로 빈도가 높은 여가활동으로 간주하여 이에 항목을 병합 처리하였다.

서비스(개인업무)의 경우는 해당 활동 정보를 가통

Table 1. Statistics of bring/get activity (unit: minute)

	minimum	maximum	mean	std. deviation
OVG	1	375	19.65	29.26
HTS	0	1,020	84.75	151.01

에서 획득할 수 없기 때문에(가통에서는 기타로 분류됨) 가통의 여러 통행 목적 중 비교적 빈도가 높고(0.8%) 연구지역의 통행 수요에 있어서 그 중요도가 높은 학원을 대신 적용하였다. 해당 작업은 서비스와 학원 활동 간 시간특성에 있어서 큰 차이가 없으며(일부 개인업무는 학원을 포함하기도 함), 활동공간에 있어서 서비스 활동에 해당하는 토지이용정보, 즉 서비스 산업의 종사자 수를 학원 종사자 수로 변경하여 적용하였다. 보다 구체적으로, 시뮬레이터에서 정의하는 활동유형(자료)은 시스템 상에서 해당 활동의 시공간적 특성과 결부되는 토지이용 및 네트워크 자료와 연결되어 있는데, 이러한 자료 변형 시 자료의 일관성을 유지하는 것이 중요하다. 활동기반모형에서 정의하는 각 활동유형은 그 활동공간 선택을 위해서 공간단위별(예로, TAZ별) 매력도/흡입력에 대한 정보가 필요하며, FEATHERS의 경우 이를 위해 각 활동별(산업영역별) 종사자 수 정보를 사용하고 있는데, 학원의 경우는 학원업계 종사자 수를 적용하므로, 기존의 서비스 활동을 학원 활동으로 변경하려면 해당 산업영역에 맞는 종사자 수 자료로 대체하게 된다.

이외에도 마중·배웅 활동에서 흥미로운 점을 발견하였는데, Table 1에서 보듯이 OVG와 가통에서 나타난 각 마중·배웅 활동의 시간 특성이 확연하게 다를 수 있다. 이 차이는 단순히 배웅 활동에 대한 피조사자의 해석상 차이로 볼 수 있지만 더 근본적으로 배웅 활동에 대한 문화적 인식의 차이로 해석할 수 있다. 다시 말해, 서양문화권에서는 일상에서 배웅 활동은 일반적으로 자녀의 등·하교를 위한 (통행) 활동으로 간주되는 반면에, 동양문화권(특히 우리나라)에서는 등·하교뿐만 아니라 “오는 사람을 맞이하는 행위”의 마중과 “떠나는 사람을 보내는 행위”의 배웅을 모두 포함한다(국립국어원 표준어대사전 참조). 따라서 OVG에서는 시간 범주에 있어서 마중·배웅 활동 자체에 소요되는 시간만 포함하는(Table 1에서 평균 20분 정도) 반면, 가통조사는 마중·배웅 활동과 엄격하게 구분할 수 없는(최소한 조사 시 구분할 수 없는) 일련의 활동(가령, 올 사람을 기다리면서 하는 기타 행위 등)이 배웅 활동에 포함되어

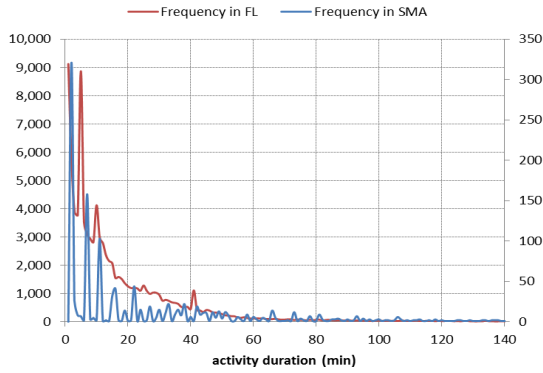


Figure 1. Frequency of bring/get activity based on starting time(unit: minute)

비교적 장시간 활동(Table 1에서 평균 약 85분)으로 보고될 여지가 있다.¹⁾

이 같은 특성을 직접 확인하기 위해서 두 조사에서 배웅 활동의 시간 분포를 보다 상세히 살펴보았다. Figure 1은 OVG("Frequency in FL")와 가통조사("Frequency in SMA")에서 나타난 마중·배웅 활동의 지속시간에 대한 빈도분포를 나타낸 것이다.

예상과 달리, 두 조사에서 활동 지속시간에 대한 전반적인 패턴이 유사한 것으로 나타났다. 참고로, 가통의 활동 지속시간 분포곡선이 OVG의 곡선보다 들쭉날쭉한 것은 가통의 피조사자에 의해 보고된 지속시간이 5분, 10분 단위로 이산성을 띄고 있기 때문이다. Figure 1에서 보듯이, 앞서 마중·배웅 활동의 지속시간특성(평균치)은 일부 피조사자에 의해서 보고된 이상치(outlier)에 의한 것으로 판단된다. FEATHERS는 이 같은 이상치에 대해서 각 활동유형의 지속시간에 대한 임계치를 설정하여 분석에서 제외시킨다. 다음은 FEATHERS의 핵심 엔진인 ALBATROSS에서 설정한 활동유형별 지속시간의(최대) 임계치 값이다.

- 업무 < 350분
- 마중·배웅 < 60분
- 기타 < 500분

위 임계치는 ALBATROSS가 개발된 네덜란드 지역에서 실시된 현장조사를 토대로 설정되었으며, 이 값은

Table 2. Activity duration (unit: minute)

activity type	observation	prediction with previous threshold	prediction with new threshold
business	219	142	180
bring/get	89	24	34
other	170	154	134

벨기에 플란더스 지역에서도 무리 없이 적용되는 것으로 나타났다. 따라서 두 국가와 여러 맥락에서 지역적 특이성이 차별되는 수도권에 대해서 기 설정된 임계치가 잘 적용되는지에 대한 사전조사가 필요하다. 조사 결과 마중·배웅의 임계치 60분을 적용할 경우에 이상치에 해당하는 활동의 빈도는 전체 마중·배웅 활동의 16.9%로 나타났으며, 이 임계치를 그대로 적용할 경우에 모의 과정에서 상당수의 자료들이 배제될 가능성이 높다. 따라서 본 연구에서는 비교적 계산이 용이하고 여러 연구에서 이상치 탐색방법으로 활용되고 있는 Devore(2000)의 Quartile or Fourth-Spread method를 이용하여 연구 지역에 적합한 임계치(UB: upper boundary)를 식(1)로 계산하였다(여기서, m은 중위수, fs는 fourth-spread, Q3, Q1은 각각 75%, 25% 사분위수).

$$UB = m + 1.5 \times fs \quad (fs = Q_3 - Q_1) \quad (1)$$

Table 2는 새로 계산된 임계치(각각 업무 500분, 마중/배웅 150분, 기타 400분)를 모의실험에 적용한 결과를 나타낸 것으로, 각 유형별로 활동지속시간의 예측 정확도 개선된 것을 확인할 수 있다.

무엇보다 조사항목을 보다 명확하게 정의함과 동시에 조사 현장에서 피조사자에게 가능한 한 상세히 설명하여 활동-통행 요소에 대한 오해의 소지가 없도록 하는 것이 위와 같은 이상치에 의한 자료 손실을 최소화할 수 있으며, 아울러 시뮬레이터에 의한 예측결과의 정확도를 개선할 수 있는 방법임은 자명한 사실이다.

한편, 가통에서 통행 수단은 승용차, 승합차, 광역버스, 간선버스, 지하철, 고속철도 등 총 18종류의 통행 수단을 구분하지만, FEATHERS는 자가용 운전자, 도보, 대중교통수단, 자가용 동승자와 같이 대분류로 설정한다. 따라서 가통의 통행 수단을 FEATHERS의 통행 수단 분류 체계에 맞추어 다음과 같이 재분류하였다.

1) OVG는 마중/배웅에 수반되는 다른 활동을 개인업무(서비스)로 따로 처리하기 때문에 활동지속시간이 짧게 나타나는데 비해, 가통은 그에 수반된 활동이 마중/배웅 활동으로부터 구분되지 않고 포함 처리되는 경우가 있어 활동지속시간이 결과적으로 길게 나타나게 된다. 이와 더불어, Table 1의 가통의 마중·배웅 시간은 FEATHERS와 마찬가지로 19세 이상의 성인의 자료만을 추려서 분석한 결과이다. 청소년을 포함한 전체 인구를 대상으로 할 경우에 약 23분으로 나타난다.

Table 3. Distribution of single-modal and multi-modal trips in survey

	single-modal	multi-modal	sum
total	1,329,165	223,55	1,351,520
	98.3%	1.7%	100.0%

- 자가용 운전자 = {모든 자가용 운전 수단}
- 도보 = {도보, 자전거, 오토바이²⁾}
- 대중교통수단 = {지하철, 기차, 버스 등}
- 자가용 동승자 = {자가용 동승자}

연구지역의 여러 대중교통 수단의 차별적인 서비스에 따른 개인의 활동-통행 의사결정 과정에 대한 다양한 영향 정도를 가늠해 볼 때 통행 모델 내부에서 수단간(특히, 대중교통 수단간) 영향 정도가 과소평가될 것으로 추정된다. 이후에 다양한 대중교통수단에 대한 영향을 모의할 수 있는 현재 시스템의 개선이 필요하다.

또한, 가통의 통행 자료는 경유지나 환승 정보를 담고 있는 다중통행수단을 가정하고 있으나, FEATHERS는 단일 통행 수단으로 하나의 대표수단만을 통행수단으로 정의한다. 따라서 다중 통행 수단의 가통자료를 FEATHERS 입력 자료로 활용하기 위해서는 각 구성 통행(segment) 중 최장거리 통행에 사용된 통행 수단을 대표 수단으로 설정하여 다중 수단 통행을 단일 수단 통행으로 변환해서 적용해야 한다. 가통자료의 다중 통행 수단을 단일 통행 수단으로 변환 시 기존의 수단별 통행서비스가 단일 (대표)통행수단의 정보로 대체되기 때문에 이 과정에서 어느 정도의 정보 손실 및 이로 인한 시뮬레이션 결과의 오류를 예상할 수 있다. 통행 수단의 단순화로 인한 영향 정도를 평가하기 위해서 먼저 전체 통행량에서 다중 통행 수단이 차지하는 비중을 살펴보았다. Table 3에서 보듯이, 가통자료의 전체 통행량 약 130만 건 중 약 22만 건(약 1.7%) 정도가 다중통행수단으로 나타났으며, 이는 전체 통행량 대비 적은 양으로 통행수단 단순화로 인한 영향이 미미할 것으로 보인다. 참고로, 연구지역에서 다중 수단 통행은 대부분 버스와 전철이 연계된 것으로 나타났다.

이어서 가통 자료의 통행 수단 정보를 단순화시키기 이전과 이후의 통행 특성을 간단한 통계치로 비교하여 보았다. Table 4에서 보듯이, 개인당 일일 평균 통행 거리는 이전과 이후가 각각 16.72km, 16.58km로 거의

Table 4. Statistics of trip distance and duration

statistics	before simplifying	after simplifying
trip distance(km)	16.72	16.58
trip duration(minute)	70.45	72.31

유사하며, 마찬가지로 일일 평균 통행 지속 시간도 각각 70.45분, 72.31분으로 거의 유사한 것으로 나타났다. 따라서 통행 수단 정보의 단순화로 인한 정보 손실 정도와 그에 따른 시뮬레이션 결과에의 영향 정도는 극히 적을 것으로 판단된다.

또한, 가통 자료에서 통행과 통행 사이를 활동으로 간주할 때, 활동 시간이 0인 자료(즉, 이전 통행의 종료 시간과 다음 통행의 시작 시간이 동일한 경우)는 실질적으로 활동 시간을 산정할 수 없기 때문에 입력 자료에서 제외되어야 하지만, 통행 자체가 목적인 경우(예로, 배웅 및 마중은 활동 자체의 지속시간이 매우 짧아 보고되지 않는 경우)도 있기 때문에 이 경우에 입력 자료를 그대로 유지하였다.

이상에서 언급한 전처리 작업 이외에 설문조사 당시에 발생하는 오류를 처리하는 문제가 남아있다. 이 같은 오류는 크게 미응답, 형식적 오류, 논리적 오류로 구분할 수 있는데, 미응답은 피조사자가 어떠한 이유든 특정 항목에 대해 응답하지 않은 정보로, 시뮬레이터의 실행 과정에서 문제를 예방하기 위해서 결측(Null)으로 처리하거나, 해당 가구 또는 가구원 전체의 자료를 적용 대상에서 제외한다. 둘째, 형식적 오류는 피조사자의 응답이 조사 형식에 맞지 않는 경우로, 가령 출생년도를 묻는 질문에 본인의 연령을 기입하는 식의 오류를 뜻한다. 이것은 잘못된 응답으로부터 추정이 가능한지의 여부에 따라 값을 수정하거나, 결측으로 처리하여 오류를 수정한다. 마지막, 논리적 오류란 응답 내용에 형식상 문제는 없지만 논리적 오류가 포함된 경우로, 가령 이전 통행 도착 시간이 다음 활동 시작 시간보다 늦는 경우, 또는 연속된 두 활동의 장소가 다름에도 통행이 보고되지 않은 등의 경우를 들 수 있다. 일반적으로 전자는 응답자들이 예상한 시간보다 통행 시간이 지체되어 통행 시간만 수정된 경우이고, 후자는 두 활동 사이의 통행 시간이 짧아(대부분 10분 이내 도보 통행) 보고되지 않은 경우에 발생한다. 따라서 전자의 경우는 이후 활동 시작 시간을 이전 통행 종료 시간으로 적용하고, 후자는 도보를 이용한 5분

2) 가통조사에서는 오토바이가 다른 도보수단과 비교해 통행시간, 거리 측면에서 유사한 패턴을 보이므로 도보로 통합하여 적용하였다. 마찬가지로 이유로 FEATHERS는 오토바이를 도보 수단으로 분류하고 있다.

내지 10분의 통행을 임의로 삽입하거나 통행정보가 누락된 해당 레코드를 삭제하여 자료를 정정하였다.

3. 인구 자료 생성

본 연구에서는 기존 FEATHERS가 IPF를 이용한 것과 달리, 또 다른 인구전수화 기법인 IPU에 2010년 가통 자료의 가구 및 가구원 정보와 수도권교통본부에서 구축한 2010년 사회경제지표를 각각 참조자료와 마진자료로 적용하여 FS의 인구 자료를 생산하였다.

Ye et al.(2009)에 의해서 제안된 IPU는 기존 IPF이 가진 가구와 가구원의 다계층 인구 정보를 동시에 추정할 수 없기 때문에 전수화 과정에서 두 계층(단위) 사이의 연결성을 유지할 수 없는 한계점을 극복하고자 개발되었다. IPU는 리스트 형식으로 된 참조자료를 그대로 피팅 과정에 적용할 수 있기 때문에 IPF의 피팅 과정을 위한 참조자료와 마진자료의 각 항목으로 구성된 다차원분할표(multi-way contingency table)를 작성할 필요가 없으며, 동시에 다차원분할표를 구성하는 각 차원(항목)의 교집합에 해당하는 참조자료의 특정 빈도가 0인 경우에 발생하는 Zero-cell 문제로부터 자유롭다. 하지만 IPU는 피팅 과정에서 가구별 전수화 가중치를 계산하기 위해 가구 및 가구원 자료의 가능한 모든 항목별 조합으로 구성된 가구·가구원 유형을 이용하기 때문에 전수화 대상의 자료 및 항목이 늘어날 경우에 해당 가구 및 가구원 유형이 많아져 자료 공간이 증가하고 처리 속도가 떨어지는 단점이 있다(Pritchard and Miller, 2012).

IPU를 적용한 인구전수화 결과는 Table 5와 6과 같다. 앞 장에서 설명했듯이, FEATHERS의 가구 단위 인구자료 항목의 대안으로써 가통 자료의 주택 형태, 주택 점유 형식, 가구원 수를 적용하여 전수인구 자료를 생산하였다. 가구원 단위의 항목은 연령대와 성별을 적용하였다. 전수인구 자료의 정확도는 SRMSE (Standardized Root Mean Square Error)를 식(2)에 의하여 평가하였다(t : observed, \hat{t} : predicted).

$$SRMSE = \frac{\sqrt{\sum (t - \hat{t})^2}}{N} / \left(\frac{\sum t}{N} \right) \quad (2)$$

Table 5는 가용한 마진자료에 의해서 직접 피팅된 전수 결과로, SRMSE의 값이 0.0029로 상당히 높은 정

Table 5. Synthetic result of controlled variables (SRMSE: 0.0029)

attributes	category	marginal	prediction
house type	apartment	48.4	48.3
	others	51.6	51.7
house ownership	own	46.4	46.6
	lease	29.5	29.6
	monthly rent	21.6	21.5
	others	2.4	2.2
age class(yr.)	0-19	23.3	23.3
	20-34	23.2	23.2
	35-54	34.6	34.7
	55-64	9.6	9.6
	65-74	6.0	6.0
	75+	3.2	3.2
gender	male	48.9	48.9
	female	51.1	51.1
nr. of members	1	21.9	21.9
	2+	78.1	78.1

Table 6. Synthetic result of uncontrolled variables (SRMSE: 0.0953)

attributes	category	seed	prediction
nr. of preschool children	0	87.1	86.0
	1	9.5	10.5
	2	3.2	3.2
	3	0.2	0.2
	4+	0.0	0.0
nr. of own vehicles	0	28.0	34.7
	1	60.1	55.2
	2	11.3	9.5
	3	0.6	0.5
	4	0.0	0.0
income level	unknown	0.5	0.5
	0-100	12.2	16.1
	100-200	20.1	23.2
	200-300	28.2	27.0
	300-500	27.1	23.3
	500-1000	10.5	8.8
	1000+	1.4	1.2
driver license ownership	no	43.5	44.7
	yes	56.5	55.3
relationship with householder	householder	34.2	36.6
	partner	26.2	24.4
	child	35.6	34.2
	parent	2.3	2.9
	others	1.6	2.0
employed	no	52.7	52.7
	yes	47.3	47.3
job	unemployed	52.7	52.7
	professional	5.7	5.9
	service	7.1	7.3
	sales	6.4	6.4
	manager	12.7	12.3
	agriculture	1.4	1.1
	labor	6.5	6.6
	others	7.5	7.7

확도를 나타낸다. 반면 Table 6은 해당 항목에 대한 마진자료의 부재로 인해 피팅 과정에 직접 참여하지 못한 전수화 결과를 나타낸 것이다. 다음 항목들은 마진정보로 이용되지 않았음에도 SRMSE 수치가 0.0953으로 높은 정확도를 보이고 있다.

이는 미적용 항목과 마진 정보로 적용된 항목의 급간별 분포의 특성 간의 높은 상관관계에서 기인한 것으로 판단된다. 다시 말해, 연구지역에서 특정 인구집단의 주택 형태, 주택점유형식, 가구원수, 연령 및 성별의 분포 특성은 가구내 아동 수, 자동차 면허소유 여부, 직업유무 등 다른 항목의 분포 특성을 잘 대변하고 있기 때문에 마진정보로 이용되지 않더라도 전수화 결과에 그 분포가 잘 예측된 것이다. 그러나 가구 내 승용차 수, 가구소득 등의 일부 세부항목들에서 관측치와 비교적 큰 오차를 나타내는데, 이에 대한 추가적인 연구가 필요하다.

4. 환경 자료: 토지이용 및 네트워크

FEATHER의 공간단위별 토지이용 특성과 통행 서비스로 구성된 환경정보는 대부분 각 항목별로 필요한 정보를 해당 자료 생산 기관에 의뢰하거나 직접 원자료를 정산하여 적용했기 때문에 별다른 전처리 과정이 없었다. 입력 자료의 항목은 아니지만 시뮬레이터 내부에서 모델링의 중요한 파라미터로 사용되는 몇몇 환경정보를 연구지역의 특성에 적합하도록 튜닝하는 작업이 요구되었다. 이 파라미터에 해당하는 것은 쇼핑센터의 개·폐점시간과 연구지역의 (일반적인) 통행침두시간대에 관한 정보이다. 쇼핑센터의 개·폐점시간 정보는 쇼핑 활동의 시간제약, 즉 시간대별 해당 활동 유형의 가능 여부를 결정하는 요소로 작용한다. 통행침두시간대 정보는 침두와 비침두 시간대의 승용차 통행 시간과 대중교통 및 승용차 통행 비용을 구분하기 위한 모델 요소이다.

먼저 플란더스 (위)와 수도권 지역 (아래)의 요일별 쇼핑센터 운영시간을 비교해 보면 다음과 같다. 참고로 -9는 휴무일을 뜻하며, 수도권의 경우 가통에서 정기와 비정기 쇼핑을 구분할 수 없기에 비정기 상점의 정보를

sector	mon1	mon2	sat1	sat2	sun1	sun2
daily	08:00	20:00	08:00	20:00	-9	-9
non-daily	13:00	18:00	9:00	17:00	-9	-9

sector	mon1	mon2	sat1	sat2	sun1	sun2
daily	9:00	24:00	9:00	24:00	9:00	24:00
non-daily	-9	-9	-9	-9	-9	-9

Table 7. Patterns of daily shopping activity

	observed	SMA	FL
activity duration(min)	113.0	167.8	163.8
trip distance(km)	13.0	25.0	16.0

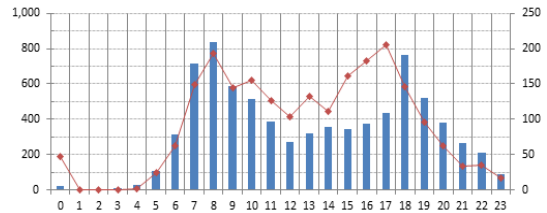


Figure 2. Frequency of trips over time (bar: SMA, line: Flanders)

모두 -9로 처리하여 적용 대상에서 제외시켰다.

전반적으로 플란더스는 주중에 오전 8시에 개점해서 오후 8시 정도에 폐점하는 반면에, 수도권은 오전 9시에 개점해서 자정쯤 폐점하여 비교적 장시간 운영하는 것으로 나타났다(참고로, 수도권의 상점 운영 시간 정보는 대형마트의 운영 시간을 기준으로 설정하였다). 따라서 쇼핑활동에 있어서 수도권에서의 시간제약이 플란더스에 비해서 덜하여 크게 시간에 구애받지 않고 쇼핑활동을 선택할 수 있다.

다음은 위 두 지역의 상점 운영 시간 정보를 각각 적용하여 쇼핑활동에 대한 시뮬레이션 결과의 정확도를 비교하였다. Table 7에서 보듯이 일평균 쇼핑활동의 지속 시간은 두 지역의 상점 운영 시간 차이와 무관하게 167분(수도권 경우)과 163분(플란더스 경우)으로 거의 비슷한 결과를 보였다. 하지만 쇼핑 목적의 일평균 통행 거리는 25km(수도권), 16km(플란더스)로 비교적 큰 차이를 보였다. 또한 관측치(가통) 13km와 비교해 볼 때 오히려 플란더스의 상점 운영 시간을 적용한 경우에 시뮬레이션 결과의 정확도가 높은 것으로 나타났다. 복잡한 메커니즘에 의해 계산된 시뮬레이션 결과를 직관적으로 해석할 수는 없지만 이는 수도권의 상점 운영 시간에 있어서 높은 자유도(낮은 시간제약 조건)로 인해서 오히려 시뮬레이션 과정에 오차를 발생시킨 것으로 추정된다. 시뮬레이션 결과에 대한 분명한 원인을 밝힘과 동시에 상점 운영 시간 정보의 영향 정도를 평가하기 위한 추가 연구가 요구된다.

다음은 플란더스와 수도권 지역의 시간대별 통행량을 확인하여 각각 침두시간대를 비교해 보았다. Figure 2는 각각 두 지역의 시간대별 통행 빈도를 나타낸 그래프이다. 오전 침두시간은 7-9시 사이로 두 지역에서 유사

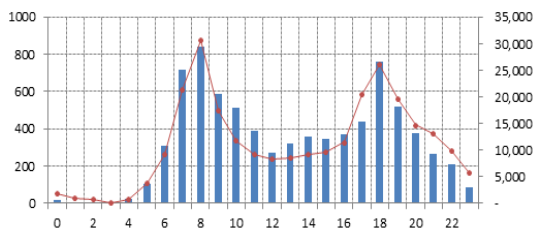


Figure 3. Frequency of trips over time (bar: observed, line: prediction)

하지만, 오후 침두시간은 플란디스에서 16-18시, 수도권에서 17-19시에 각각 다르게 분포한다. 따라서 연구 지역인 수도권의 통행 특성에 맞도록 오후 침두시간 정보를 17-19시로 수정하였다.

침두시간 변경 이후에 시간대별 통행량의 관측치(가통 자료)와 시뮬레이션 결과인 추정치를 비교하여 정확도를 분석하였다. Figure 3은 시간대별 관측치와 추정치의 통행량을 나타낸 것이다. 분석 결과 관측치와 매우 유사한 패턴을 보여 높은 예측 정확도를 나타냈다. 반면에 오후 침두시간대 정보의 전후를 비교해 본 결과 해당 정보 변경으로 인한 시뮬레이션 결과의 영향 정도는 미미한 것으로 나타났다. 하지만 추후 통행량뿐만 아니라 좀 더 다양한 분석을 통해서 모델에 대한 침두시간 정보의 영향을 평가할 필요가 있다.

결론 및 토의

이 논문은 FEATHERS 활동-통행 시뮬레이터를 수도권 지역에 적용하기 위한 FS 개발에 있어서 입력자료 전처리과정에서 발생하는 각종 문제에 대한 개선 방안을 제시하고, 그 효과를 직접 실험을 통해 검증하였다. 먼저, FEATHERS와 입력 자료로 사용된 가통의 자료구조 상 불일치와 구성요소의 해석상 차이로 인해 발생하는 문제에 대해 논의하고, 그 해결방안을 제시하였다. 또한 실제 적용 결과, 다이어리 자료의 정확도 개선에 어느 정도 기여하였으며, 전수인구 자료 생성에 있어서 마진 정보의 적용 유무와 상관없이 관측치와 매우 일치하여 높은 예측력을 나타내었다. 따라서 본 연구에서 제안한 자료 전처리 개선방안은 분명한 성과를 거두었음을 알 수 있다.

이상의 성과에도 불구하고 이 연구는 다음과 같은 향후 연구 과제를 확인할 수 있다. 첫째, 센서스에 마진 정보가 없는 변수들에 대해서도 100%에 가까운 전수화

정확도를 얻어야 한다. 물론 마진 정보가 있는 변수들이 그렇지 않은 변수들과의 반복적인 정산 과정에도 불구하고 100% 가까운 정확성을 보이고 있음은 다행스런 결과이다. 그러나 마진 정보가 없는 변수들에 대해서도 이와 유사한 수준의 결과를 얻어야 활동-통행 시뮬레이션 결과가 신뢰성을 확보할 수 있을 것이다. 이에 대한 추가적인 이론적 연구가 필요하나, 보다 근본적인 해결책은 센서스가 중요 변수들에 대한 가구 마진 정보를 제공하도록 수정되는 것이다.

둘째, 카테고리 불일치의 문제에 대한 보다 근본적인 해결책 모색이다. 통행 수단 항목의 불일치, 활동 유형 항목의 불일치 등은 시뮬레이터의 예측성을 크게 떨어뜨린다. 이의 근본적인 해결책은 아마도 우리 자체적인 소스코드의 개발일 것이다.

셋째 대중교통수단 간 연계(버스와 지하철 간 환승) 및 동일수단 간 연계(버스노선 간 환승) 등의 다중수단은 가통에서 차지하는 비중이 비록 적으나, 이를 하나의 단일수단으로 처리한 것은 이 연구의 분명한 한계이다. 수단 선택의 정확한 분석은 물론 관련된 정책의 중요성을 고려할 때 시스템에 반드시 추가 보완해야 하는 요소이다.

넷째 활동 유형에 대한 보다 근본적인 성찰이 필요하다(Doherty, 2003). 연구자 임의로 정하는 활동 목록은 연구 결과의 신뢰성에 의문을 던지게 한다. 명목적으로 주어진 불변의 활동 유형 이름은 사실은 그와 다른 내용을 수행하는 것일 수도 있다. 예로 점심 식사를 식사로 볼 것인가 아니면 그 시간에 이루어지는 중요한 비즈니스 미팅으로 볼 것인가와 같은 문제로서, 이의 결정은 해당 사회의 문화적 내용이나 활동 실행 당시의 상황에 따른다고 보인다.

다섯째 이와 같은 입력자료 구조 불일치 문제를 해결 하더라도 남은 문제로서, 시뮬레이터가 개발된 지역의 사회경제문화적 특성을 반영한 의사결정 모듈을 갖는 것이다. 이는 즉 스케줄링 엔진을 궁극적으로 우리 수도권 주민이 갖는 의사결정 방식을 갖는 것으로 바꾸는 것을 뜻하는데, 이를 위해 수도권 주민을 위한 별도의 SP(stated preference)조사나 활동-통행 의사결정에 관한 실증연구가 요구된다.

이 연구는 시뮬레이터의 실행 측면 보다는 시뮬레이터 실행을 위한 준비 과정, 즉 전처리 과정과 입력자료로 사용할 가통 조사에 좀 더 포커스를 두었다. 따라서 다른 지역, 다른 환경에서 개발된 모델을 전용함에 있어서 수

도권의 통행기반 조사 자료가 얼마나 잘 활용될 수 있는지, 모델이 처음 개발된 지역과 전혀 다른 문화권에서 수집된 조사 자료가 활동기반 측면에서 어떤 문제점을 가지고 있는지, 그 해결 방안은 무엇인지 등에 더 큰 관심을 가졌다. 이에 따라 연구지역에 대한 보다 상세한 시뮬레이션 연구가 필요한데, 이는 이번 연구를 토대로 실행 측면을 고려한 향후 연구에서 중점적으로 다루어야 할 일이다.

ACKNOWLEDGEMENT

This work was supported by the National Research Foundation of Korea grant funded by the Korea Government(MSIP) (NRF-2010-0029444).

REFERENCES

- Arentze T.A., Hofman F., van Mourik H., Timmermans H.J.P. (2000), Albatross: Multi-agent rule based model of activity pattern decisions, *Transportation Research Record*, 1706, 136-144.
- Arentze T.A., Hofman F., van Mourik H., Timmermans H.J.P. (2002), Spatial Transferability of the Albatross Model System: Empirical Evidence from Two Case Studies, *Transportation Research Record: Journal of the Transportation Research Board*, 1805, 1-7.
- Arentze T.A., Timmermans H.J.P. (2012), Travel demand modelling: Conceptual developments and perspectives, *Transportation Letters*, 4, 85-92.
- Bellemans T., Kochan B., Janssens D., Wets G., Arentze T., Timmermans H.J.P. (2010), Implementation framework and development trajectory of FEATHERS activity-based simulation platform, *Transportation Research Record*, 2175, 111-119.
- Bhat C.R., Guo J.Y., Srinivasan S., Sivakumar A. (2004), A Comprehensive Econometric Micro-simulator for Daily Activity-travel Patterns (CEMDAP), Paper presented at the 83rd Annual Meeting of TRB, Washington D.C.
- Devore J.L. (2000), *Probability and Statistics for Engineering and the Sciences*, 5th ed. Duxbury, Pacific Grove, CA.
- Doherty S.T. (2003), Should we abandon activity type analysis? Paper presented at the 10th International Conference on Travel Behavior Research, Lucerne.
- Goulias K.G., Bhat C.R., Pendyala R.M., Chen Y., Paleti R., Konduri K.C., Lei T., Tang D., Yoon S.Y., Huang G., Hu H.H. (2012), Simulator of Activities, Greenhouse Emissions, Networks and Travel (SimAGENT) in Southern California, Paper presented at the 91st Annual Meeting of TRB, Washington D.C.
- Guo J.Y., Bhat C.R. (2007), Population synthesis for microsimulating travel behavior, *Transportation Research Record*, 2014, 92-101.
- Knapen L., Kochan B., Bellemans T., Janssens D., Wets G. (2012), Using activity-based modeling to predict spatial and temporal power demand of electric vehicles in Flanders, Belgium, *Transportation Research Record*, 2287, 146-154.
- Kochan B., Bellemans T., Cools M., Janssens D., Wets G. (2011), An estimation of total vehicle travel reduction in the case of telecommuting, Detailed analyses using an activity-based modeling approach, Paper presented at European Transportation Conference, Glasgow.
- Lee B.J., Lee C.Y., Kim J.K. (2011), Mobility management for sustainable urban transportation policy paradigm (지속가능한 도시교통정책 패러다임 모빌리티 매니지먼트), The 64th Conference of Korean Society of Transportation, Korean Society of Transportation, 177-193. (in Korean)
- Lee W.D., Joh C.H., Cho S., Kochan B. (2014), Issues in Feathers application in the Seoul Metropolitan Area, *Data Science and Simulation in Transportation Research*, IGI, 71-85.
- Lim G.G., Kim S.G., Jung S.B. (2013), Activity-based approaches for travel demand modeling: Reviews on developments and implementations, *Journal of the Korean Society of Civil Engineers*, 33, 719-727. (in Korean)
- Park J.Y., Lee J.S., Kim Y.H., You J.B. (2012), A Fundamental Study for Analyzing Future Human

- Mobility (2012-24), The Korea Transport Institute.
(in Korean)
- Pinjari A.R., Bhat C.R. (2010), Activity-based travel demand analysis, in A. de Palma, R. Lindsey, E. Quinet and R. Vickerman (eds.), Handbook of Transport Economics, Chapter 10, Edward Elgar Publishing, 213-248.
- Pirdavani A., Brijs T., Bellemans T., Kochan B., Wets G. (2012), Application of different exposure measures in developing planning-level zonal crash prediction models, Transportation Research Record, 2280, 145-153.
- Pritchard D.R., Miller E.J. (2012), Advances in Population Synthesis: Fitting Many Attributes per Agent and Fitting to Household and Person Margins Simultaneously, Transportation, 39(3), 685-704.
- Sikder S., Pinjari A.R., Srinivasan S., Nowrouzian R. (2013), Spatial Transferability of Travel Forecasting Models: A Review and Synthesis, International Journal of Advances in Engineering Sciences and Applied Mathematics, 5(2-3), 104-128.
- Timmermans H.J.P., Arentze T.A., Joh C.H. (2002), Analysing space-time behaviour: New approaches to old problems, Progress in Human Geography, 26, 175-190.
- Ye X., Konduri K.C., Pendyala R.M., Sana B., Waddell B. (2009), A methodology to match distributions of both household and person attributes in the generation of synthetic populations, Paper presented at the 88th Annual Meeting of TRB, Washington D.C.

- ☞ 주 작성자 : 조성진
☞ 교신저자 : 조창현
☞ 논문투고일 : 2014. 7. 23
☞ 논문심사일 : 2014. 9. 27 (1차)
2014. 10. 14 (2차)
☞ 심사판정일 : 2014. 10. 14
☞ 반론접수기한 : 2015. 2. 28
☞ 3인 익명 심사필
☞ 1인 abstract 교정필