

AI 그리고 온디맨드 교통정책

2018년 예산안의 핵심 기조는 새 정부의 성격을 잘 드러내고 있다.¹⁾ 새해 예산안의 골자는 큰 폭의 사회간접자본(social overhead capital, SOC) 예산 삭감과 큰 폭의 복지 확대이다. 이는 소득 주도 성장을 구체화하려는 새 정부의 지향성과 변화된 시대 정신이 반영된 결과다. 경기부양을 위해 비교적 단기간(예를 들어, 선출직 공무원의 치적으로 산입이 가능한 기간) 안에 효과를 기대할 수 있는 건설 경기 부양 카드를 꺼내 부동산을 축추게 함으로써, 결과적으로 국가 경제 체질이 악화되는 악순환의 고리는 이제 정리되는 듯 싶다. 정부의 최근 예산 기조를 살펴 보면, 국가 경영의 중점이 물적 기반에서 인적 자원으로, 총량적 성장에서 개인 복지와 경제 활동의 지속가능성을 강화하는 방향으로 바뀌었다.

교통정책의 변화: 공급 위주에서 온디맨드 정책으로

공급 위주의 교통 정책이 개별 교통 수요자 중심, 즉 온디맨드 교통 정책으로 변화하는 것은 개발 시대에서 성숙한 사회로 나아가는 시대정신 변화의 사례다. 이는 자율주행, 사물인터넷 등으로 대표되는 최근 교통관련 기술의 발전, 시민사회 중심의 도시 경영 등이 동인으로 작용한 결과다. 그러나, 교통과 관련한 보다 직접적이고 원천적인 변화의 동인은 교통 현상 자체를 기준으로 한 교통 공급의 최적화와 파생 수요로서의 교통 수요 관리의 최적화를 교통 정책의 핵심으로 간주하는 인식의 변화다.

인식의 변화는 새로운 시대의 교통 계획에서 보이는 특징이다. 새 시대의 교통 계획은 교통 현상이 교통 주체의 일상 생활(여객) 및 경제 행위(화물)의 수행을 위해 불가피하게 파생된다는 점에 주목한다. 그리고 교통 주체의 일상 생활과 경제 행위를 이해하는 것을 매우 중요시한다. 이는 교통을 바라보는 시각의 일대 전환으로 매우 중요한 시사점을 갖는다.

문제는 교통 주체의 일상 생활과 경제 행위를 이해하기 위해서는 기존의 교통 현상 이해를 위한 일반적인 연구 범위를 넘어서는, 사회공간적이고 경제공간적인 고찰이 필요하다는 것이다. 개인이 어떻게 일상을 조직하는가, 개인이 요일에 따라, 상황에 따라 어떻게 일상을 변화시키는가에 따라 파생되는 교통 수요의 내용은 달라진다.

예를 들어 지정된 시공간으로 규칙적인 출퇴근을 하는 사무실 근로자와 주문에 따라 행선지가 바뀌는 영업사원은 하루 통행의 양상과 형태가 매우 다르다. 사무실 근로자 역시 승용차가 있는지 없는 지에 따라, 또 주중과 주말에 따라 이동의 양상이 다르며, 주말의 통행을 가족과 하느냐 친구와 하느냐에 따라 통행 내용은 또 달라진다. 심지어 기상 상태에 따라, 교통 현상은 많은 변화를 보인다.

당연하게도, 개인 교통 수요의 변화는 그 지역 사회의 집합적 교통 수요 변화로 이어진다. 다시 말해 서울시, 종로구, 혹은 혜화동 등 공간적 스케일과 하루의 시간대와 같은 요인에 따라 개인의 교통 수요는 모두 다르지만, 개인들이 갖는 이러한 다양한 요인에 의한 통행 양상과 형태가 집합적으로 발현됨으로써 교통 현상이 관찰되는 것이다.

이러한 시각은 교통 혹은 비(非)교통 정책이 사람의 일상을 바꾸고 이것이 다시 연쇄적으로 통행 행동을 바꾸는 일반적인 경우에 해당된다. 지난 수 년 간 사회적 이슈가 되어 온 경전철, 공항 철도 등 대규모 교통 시설의 공급 후 수요가 예상치에 크게 미치지 못하여 대규모의 적자가 발생하는 사태²⁾³⁾는 이러한 사회공간적 변화와 교통 현상 간의 연쇄 상호작용을 연구할 필요를 잘 보여준다.

유럽의 민주적 도시계획 전통: 파생 수요로서의 교통량 시뮬레이션

파생 수요로서의 교통 현상의 변화가 개인 단위로 일어나고 이를 집합적으로 관찰하여 교통 정책에 반영하는 것은 북유럽의 민주적 도시계획 전통인 파생수요로서의 교통량 시뮬레이션의 기본 정신이다. 이는 토르스텐 해거스트란트(Torsten Hägerstrand)⁴⁾실험 성공에 기초한 것이다. 해거스트란트는 1970년대와 1980년대 전후 스웨덴의 성공적인 공간 재구조화와 도시체계 재편에 지대한 영향을 끼쳤던 도시지리학자다.

민주적 도시계획에서는, 도시의 전체 생산성이나 효율성 개선보다는 도시를 사는 주민 개개인의 삶의 질 향상을 최우선의 목적으로 인식한다. 해거스트란트가 주창한 시간지리학에서 개인 일상의 시공간 궤적은 그가 속한 사회와 그가 맺고 있는 특별한 사회공간적 관계가 표출된 결과이다. 해거스트란트는 미시적인(micro) 통행 활동을 시뮬레이션을 통해 거시적인(macro) 집합적 경향으로 읽어 내면, 다양한 정책 목표 달성을 위해 도시 공간을 어떻게 조직할 것인지, 그 안의 주민들의 삶의 질은 어떻게 향상시킬 것인지에 대한 해답을 얻을 수 있음을 이론화하였다. 그의 이론은 이후 영국의 사회학자 기든스(Anthony Giddens)에 의해 구조화 이론으로 재구성되기도 하였다.

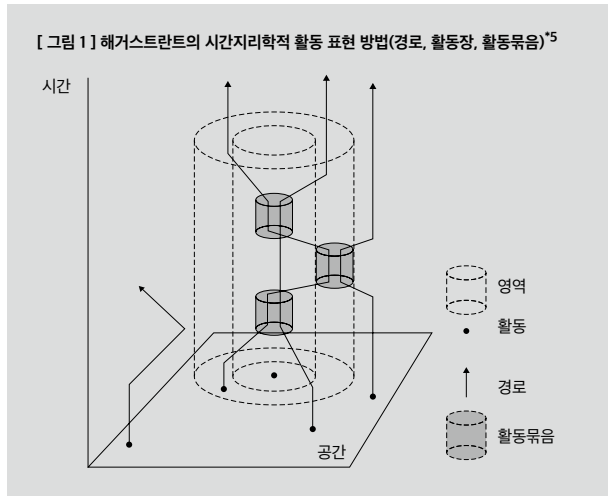
시뮬레이션의 스키마: 통행 주체별 이동 특성

미시적인 통행 활동을 시뮬레이션하는 시스템이 정의하는 시뮬레이션의 스키마(schema)는 통행 주체별 이동 특성이다. 여객 통행의 개인 이동 특성은 구성적(compositional) 방법과 맥락적(contextual) 방법으로 표현된다. 통행을 표현하는 요소에는 통행 목적, 목적지, 교통 수단, 동행인, 통행 거리 및 비용 등이 있다.

구성적 방법은 통행 횟수, 통행 거리, 동행인 수, 교통 수단별 이용 수 등으로 하루나 일주일 등 주어진 기간 동안 개인의 통행을 표현한다. 맥락적 방법은 [그림 1]에서 보듯이 해거스트란트의 시간지리학적 활동 표현 방법을 따른다. 이 방법은 하나 하나의 통행 목적에 함께 존재하는 목적지, 교통 수단 등의 정보를 하나로 묶고 나서 이를 통행 목적들 간의 시간적 선후 관계로 연결하여 하나의 궤적으로 표현한다.

글 | 조창현 biertje@daum.net

서울대, 네덜란드 아인트호벤 공대에서 수학했습니다 영국 임페리얼대에서 연구개발 했으며, 경희대에서 교통 강의를 하고 있습니다. 교통 자료의 분석 과정과 결과에 사회과학적 해석과 의미를 부여하는 일을 수행 중입니다.



이렇게 하면, 대중 교통을 타고 출근한 사람이 퇴근할 때 자신의 승용차를 이용할 수도 있다는 비논리적이고 불필요한 대안을 배제할 수 있고, 퇴근 후 쇼핑한 사람의 다음 이동은 귀가 목적의 이동일 가능성이 높다는 것 등을 기대할 수 있다. 이러한 궤적을 통행 패턴이라 하며, 유사한 통행 패턴을 가진 통행 주체들은 해당 통행 패턴의 특성과 연관된 속성들을 가진 것으로 기대될 수 있다. 유사 통행 패턴, 그리고 패턴별 통행 주체의 인구통계학적 특성, 해당 지역의 지리적 특성, 통행 당시의 상황적 특성 등은 교통 정책 수립에 주요 정보가 된다.

AI의 다차원 패턴 인식: 분자생물학적 방법과

사회과학적 해석의 만남

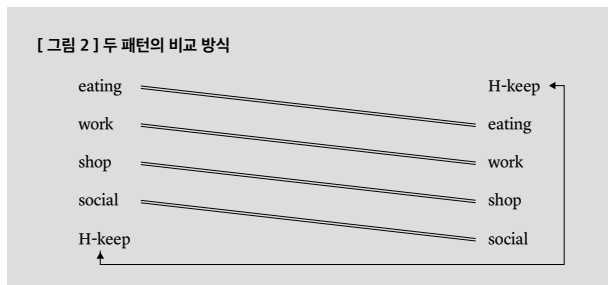
맥락적 방법으로 정의된 통행 패턴을 분류하는 방법을 만들어 내는 일은 직관적으로 어렵다. 일반적으로, 구성적 방법으로 수집된 자료는 상용 통계 프로그램을 이용하여 사례의 분류를 쉽게 수행할 수 있다. 전형적인 자료 구조는 엑셀의 작업 시트와 같이, 행은 케이스, 열은 각 통행 요소 변수가 해당 케이스를 설명하는 식으로 배열될 수 있다. 예를 들어, 하루 동안의 통행을 구성적으로 기록한 통행 패턴은 총 통행 몇 회, 승용차 이용 통행 몇 회, 총 통행 거리 몇 회 등과 같이, 개별 통행이 아니라 해당 일자의 통행 패턴에 속한 모든 통행을 요약하는 정보로 통행 패턴을 묘사한다. 상용 통계 프로그램의 군집 분석은 이 같은 내용에 대한 통행 패턴 간 유사성 정도를 측정할 거리 행렬을 만들어 정해진 군집화 알고리즘에 따라 유사한 사례들을 집단으로 분류해 낸다. 그러나 우리가 이야기하고 있는 통행 패턴은 궤적화되어 있는 맥락적 구조이다. 상용 통계 프로그램은 이러한 자료 구조를 갖는 통행 패턴을 분류할 수 있는 방법을 제공하지 않는다.

이 문제를 풀기 위해 일단 문제를 단순화해 본다. 통행 패턴의

여러 요소 중 한 가지 요소만 고려한다면, 예를 들어 통행 목적만 고려한다면, 통행 패턴은 하루 동안 한 사람의 통행 목적의 궤적으로 정의할 수 있다. 맥락적 두 패턴의 유사성과 차이점을 확인하는 방법은 두 패턴이 갖는 구성적 정보(통행 수, 목적별 통행 수)와 맥락적 정보(목적 간 순서) 모두를 한 번에 비교하는 것이다.

캐나다 주민의 시간 사용 현황을 연구하던 윌슨(W.C.

Wilson)⁶은 각 지역의 주민이 하루 동안 보낸 활동의 궤적을 비교하는 방법을 분자생물학으로부터 도입했다. 정보배열비교법(sequence alignment methods, SAM)이라 불리는 이 방법은 구성과 맥락이란 두 가지 종류의 정보를 비교하는 일을 한 번에 처리할 수 있다.⁷ 본래 이 방법은 분자생물학에서 서로 다른 생물 종 간 유사성과 차이의 정도를 측정하려는 목적으로 개발된 동적 최적화(dynamic optimization) 방법이다. 이 방법은 DNA나 RNA의 염기 배열을 비교하는 방식으로, 높은 일치도를 보이면 같은 종이라는 것을 확인할 수 있다. 또한 이 방법을 통해 사람과 침팬지 사이의 사람과 악어 사이보다 더 가깝다는 등 서로 다른 종들의 차이의 정도도 비교할 수 있다. [그림 2]는 SAM이 두 패턴을 비교하는 방식을 도식적으로 보여준다.

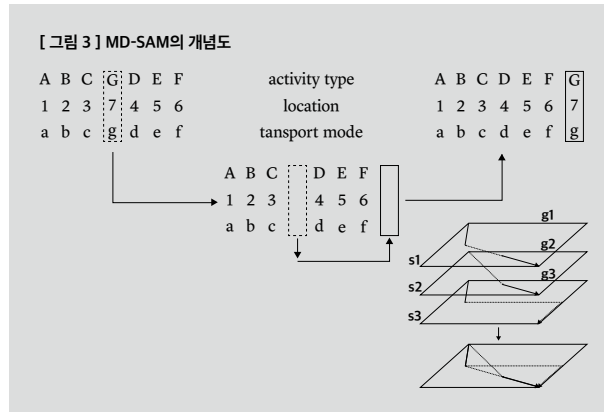


다차원 SAM

문제는 통행 패턴이 갖는 정보는 통행 목적뿐 아니라 통행 수단, 목적지, 동반자 등 다차원이며, 패턴 간 비교는 이러한 다차원의 구성 및 맥락적 정보를 다루어야 한다는 데 있다. 교통 연구에서 원래의 SAM을 확장한 다차원 SAM(multi-dimensional SAMs, MD-SAM)이 개발됐다.⁸ 이는 하나의 통행을 위해 내려지는 여러 차원의 통행 의사 결정이 한 번에 묶음으로 행해지는 것에 착안하여 두 통행 패턴 간 비교를 수행하는 방법이다.

MD-SAM은 AI에서 광범위하게 응용되는 다차원

패턴인식의 한 방법이다. 필자는 2001년에 발표했던 논문⁹에서 통행 패턴 분류의 경우 다른 대표적 패턴 인식 방법인 Walsh-Hadamard transformation의 전기전자적 방법에 비해 군집 패턴 프로파일(profile)의 명확성이 더 높음을 보여주었다. [그림 3]은 다차원 패턴 비교 방법의 개념을 모식적으로 보여준다.



원래의 단차원 SAM을 다차원으로 확장한 문제의 해를 찾는 방법은 시뮬레이티드 어닐링(simulated annealing), 신경망(neural network), 유전 알고리즘(genetic algorithm) 등 여러가지 AI적 패턴인식 방법들이 가능하나, MD-SAM에서 실용화된 방법은 재생산(reproduction), 교차(crossover), 돌연변이(mutation) 등의 수단을 활용한 유전 알고리즘을 채택하였다. 진화 이론에 근거한 유전 알고리즘은 돌연변이에 의한 종의 진화를 불규칙한 해 공간에서의 해 찾기에 응용한 방법으로 다음과 같이 설명할 수 있다.

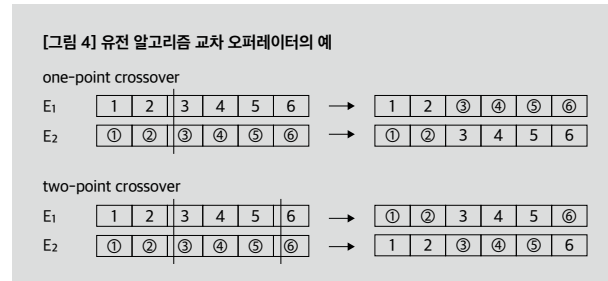
유전 알고리즘의 작동 방식

begin
 initialize organisms (1)
 evaluate initial organisms (2)
while not (stop condition) **do** (3)
begin
 select organisms (4)
 select a genetic operator (5)
 generate new organisms by applying a genetic operator (6)
 evaluate newly generated organisms (7)
end
end

- (1) 일정 개수의 초기 해를 만들어낸다.
- (2) 이들 해 각각을 평가한다. 우리의 경우 각 해는 두 패턴의 유사성이나 차이의 정도를 측정하는 것이다. 유사성이 작다면 차이가 크다는 것이다. 차이를 측정하는 방법은 여러 가지가 있고 그에 따라 차이의 결과값도 여러 가지이다. 알고리즘은 그 중 가장 효율적으로 측정할 최소치를 두 패턴 간 차이로 간주한다. 그러나 그 최소치는 불규칙한 해 공간에 있는 것으로 생각된다. 따라서 알고리즘은 초기 해를 진화적 방법으로 개선하여 최종적으로 최적의 해를 만들어내도록 한다.
- (4) 이 개선은 우수한(즉 차이 측정치가 작은) 대안들을 확률적으로 선택하는 데서 출발한다. 무조건 가장 작은 측정치의 대안을 선택하지 않고 확률적으로 선택하는 것은, 해 찾기가 국지적 해 공간에 고착되는 것을 방지한다.
- (5) 선택한 대안들을 다음 단계의 해 찾기에 전달할 방법을 선택한다. 이 방법들은 진화 이론에 근거한 것으로, 아래 문단에서 자세히 설명한다.
- (6) 선택한 방법을 선택한 대안에 적용하여 다음 단계의 해들을 만들어낸다.
- (7) 새롭게 만들어낸 해들 각각을 평가한다.
- (3) 평가한 결과가 일정 조건을 만족했는지 검토하고 만족할 만한 결과이면 그 시점의 대안을 최적이라 간주한다. 이 때 조건은 일반적으로 해 값에 더 이상의 개선이 없는지 여부로 정의된다.

유전 알고리즘은 해 공간이 불규칙할 때 최적해를 찾기 위한 더 나은 해의 탐색(exploration)과 기존 해 찾기 경험의 이용(exploitation)을 적절히 조합함으로써 국지적 최적에 빠지지 않으면서도 전역적 해의 질을 향상시키는 수치해석적 방법이다. 패턴 인식의 유사성과 차이점을 측정해 내는 정보를 제공하는 데 활용된다.

알고리즘의 작동에 투입되는 오퍼레이터는 생물 진화의 세 가지 진화 요소를 대입한다. 재생산은 조상의 우수한 유전 정보를 후대에 그대로 전달하는 요소를 흉내 내어, 현 시점의 해 찾기의 결과 중 좋은 결과를 탐색(exploit)하여 다음 시점으로 그대로 전달하는 오퍼레이터다. 교차는 조상의 우수한 유전정보들을 섞는 방식으로 이용(exploit)하여 더 좋은 결과를 시도하는 오퍼레이터다([그림 4]). 돌연변이는 조상의 더 나은 유전 정보를 탐색(explore)해 가는 해 찾기 작업이 국지적 최적에 빠지는 위험을 극복하기 위하여 현 시점의 이용(exploit) 결과와는 전혀 다른 새로운 대안을 시도하는 오퍼레이터다. 돌연변이의 확률이 너무 높으면 이용(exploitation)이 부실해 최종 해가 부실해질 수 있어서, 재생산이나 교차에 비해 매우 적은 확률로 적용되지만, 국지적 해의 위험을 줄이는 데 큰 기여를 한다.



유전 알고리즘을 탑재한 다차원 SAM은 단차원 SAM의 작동 방법과 일관되고, 또한 아래에서 설명되는 바와 같이 일상 활동 패턴 간의 차이를 설명하는 방식에 일치하기 때문에 분석의 결과를 해석하는 데 용이하다. 이는 통행 패턴 유사성 측정이라는 특정 도메인에서 유전 알고리즘이 neural network에 대해 갖는 확연한 이점이다. 이 같은 통행 패턴의 시장세분화(market segmentation)는, MD-SAM이 다차원 통행 의사 결정의 패턴 인식을 통해 유사성을 갖는 집단들을 구분해 내고, 또 각 집단의 통행 특성 프로파일(profile)과 해당 통행 주체의 사회인구학적 특성, 혹은 해당 통행 패턴 관련 지역의 지리적 특성, 혹은 해당 통행 패턴 관련 상황적 특성 간의 통계적 연관성을 확인하는 방식으로 이루어진다.

분자생물학적 패턴 인식의 사회과학적 근거

여기서 흥미로운 점은, 이러한 분자생물학적 다차원 패턴 인식 방법이 사회과학적 근거를 갖는다는 사실이다. SAM은 두 패턴 간 유사성과 차이의 정도를 두 패턴 간 거리로 간주한다. 거리는 두 패턴을 일치시키기 위해 한 패턴을 다른 한 패턴으로 변환시키는 데 드는 최소한의 노력으로 정의한다.

이 노력의 정도는 통행 패턴 요소의 삽입(insertion), 삭제(deletion), 교체(substitution)의 양으로 측정한다. 즉 비교하는 두 패턴으로 구성된 패턴 한 쌍에서, 두 패턴을 일치시키기 위해 한 패턴을 다른 한 패턴으로 변환하는 데 적은 수의 삽입, 삭제, 교체만이 필요하다면, 이 패턴 쌍의 패턴들은 많은 수의 삽입, 교체, 삭제가 필요한 다른 쌍의 패턴들보다 서로 더 유사하다고 간주할 수 있다. 생물 중에서 삽입, 교체, 삭제는 돌연변이에 의한 염기의 변화를 원 위치 시키는 데 필요한 작업이다. 즉 A라는 종이 한 번의 돌연변이로 B라는 종이 된다면, 열 번의 돌연변이가 필요한 C라는 종보다, A는 B와 확률적으로 더 유사하다는 것이다.

이러한 논리는 통행 패턴에 적용 가능하다. A라는 통행 패턴을 B 패턴으로 바꾸는 데 삽입, 삭제, 교체 중 한 번의 작업만이 필요한 데 비해 C 패턴으로 바꾸려면 열 번의 작업이 필요하다면, A는 C보다는 B와 더 유사하다는 결론을 얻을 수 있다. 앞에서 얘기한 바를 반복하면, 하나의 패턴은 통행 주체의 인구학적 특성이나 지역의 지리적 특성, 혹은 통행 당시의 상황적 특성을 반영한다.

따라서 유사한 패턴은 유사한 통행 주체의 특성이나 지역의 지리적 특성이나 상황적 특성으로부터 결과된 것이라 생각될 수 있다. 이에 비해 변환하는 데 많은 작업이 필요한 패턴은 이러한 특성들을 다룰 때 많은 돌연변이를 필요로 하며, 돌연변이가 많이 필요한 두 패턴은 유사하다고 말하기에는 확률적으로 매우 어렵다. 다시 말해 상이한 패턴이다 라고 결론 지을 수 있다. 이 모든 논의가, 통행 수요는 파생수요의 성격을 갖는다는 교통 현상에 대한 새로운 시각을 매우 잘 반영하는 것이며, 이러한 원리에 따른 통행 패턴의 분화(market segmentation)은 온디맨드 교통 정책에 중요한 정보를 제공할 수 있다.

새로운 패턴 분석의 선결 조건: 빅데이터와 AI 분석 기법

이상의 AI 기반 패턴인식 방법은 교통 현상을 분석하는 AI 기반 여러 방법 중 하나의 예이다. 통행 주체의 교통 행태를 분석하고 예측하는 데는 방대한 양의 자료가 필요하다. 따라서 교통 수요 분석 및 예측은 빅데이터 마이닝(big-data mining)과 다양한 AI 기법이 매우 필요한 분야이다.

일상의 통행 활동을 연구하는 데는 우선 활동 결과를

분석하고 해석하는 작업이 필요하다. 이 때, 위의 논의와 같은 방식의 AI 기반 패턴 인식 기법이 요구된다.¹⁰ 이에 더해 통행 주체의 의사 결정 구조를 확인하고 그에 따른 예측 시뮬레이션이 가능하려면 현재의 의사 결정 구조와 그 진화를 분석하는 AI를 활용한 모델링 기법이 필요하다.¹¹

우리나라에서의 빅데이터 및 AI 활용 교통 수요 분석 및 예측의 선도 사례로서, 한국철도연구원이 개발 중에 있는 ABATA (Activity BAsed Traveler Analyzer)¹²통행 수요 예측 모델을 들 수 있다. 이 모델은 통신 자료를 활용한 빅데이터 분석 기법의 온디맨드 교통 정책을 지향하고 있다. 향후 이 모델의 개발에 있어 중요 연구 주제는 개별 통행 주체의 의사 결정 구조를 시화 하여 모델 내에서 구현하는 것이며, 의사 결정 구조의 시화를 위해서는 통행패턴의 패턴인식 분류가 선행되어야 한다. 이 모델이 실전에 활용된다면, 예상 교통수요의 네트워크 로딩을 통해 하루 24시간 원하는 시간대의 타임 슬라이스(time slice)에서의 교통수요를 추정하고, 지역 내 교통 흐름의 공간분석이 가능할 것으로 보인다.

기존의 가구 통행 조사나 스마트 카드¹³ 자료를 넘어서, 엄청난 양의 내비게이션 자료와 통신 자료가 교통 분석을 위해 가용되고 있다.¹⁴ 교통 빅데이터의 마이닝과 AI 모델 개발은 이미 교통 연구의 대세가 되어 있다. 그 구체적인 플랫폼의 한 예는 현재 파생수요로서의 교통수요 예측을 강조하는 국내의 모델들이다. 관련된 예로는 위의 ABATA 외에, ACTOR(ACTivity-based micro simulaTOR)¹⁵, FS(Feathers Seoul)¹⁶ 등이 있다. 미래 사회의 공공서비스 수요는 민간 상품 수요와 크게 다를 것 같지 않다. 적재적소에서 주민이 원하는 서비스를 적절히 공급하는 것은 공공서비스 생존의 유일한 방안이 될 것이다. 그런 의미에서, 앞으로 빅데이터 분석 및 AI에 기반한 통행 수요 연구 분야의 급속한 발전이 기대된다.

¹⁰ 참고 | http://biz.chosun.com/site/data/html_dir/2017/08/29/2017082900452.html ¹² 참고 | http://biz.chosun.com/site/data/html_dir/2017/09/29/2017092902099.html ¹³ 참고 | <http://news.hankyung.com/article/2017060654031?nv=o> ¹⁴ 논문 | Öberg, S., (2005). Hägerstrand and the remaking of Sweden, *Progress in Human Geography*, 29 (3), 340-349. ¹⁵ 참고 | 조창현. (2013). 도시 일상생활 연구의 시공간적 접근: 활동기반 이론에 의한 통행 행태 연구의 확장, 푸른길. ¹⁶ 논문 | Wilson, W.C., (1998). Activity pattern analysis by means of sequence-alignment methods, *Environment and planning A*, 30 (6), 1017-1038. ¹⁷ 논문 | Kruskal, J.B., (1983). An overview of sequence comparison: time warps, string edits, and macromolecules, *SIAM Rev.*, 25, 201-237. ¹⁸ 논문 | Joh, C., Arentze, T., Hofman, F., & Timmermans, H. (2002). Activity-travel pattern similarity: A multidimensional alignment method, *Transportation research B*, 36, 385-403. ¹⁹ 논문 | Joh, C., Arentze, T., Hofman, F., & Timmermans, H. (2002). Activity pattern similarity: a multidimensional sequence alignment method. *Transportation Research Part B: Methodological*, 36(5), 385-403. doi:10.1016/s0191-2615(01)00009-1 ¹⁰ 논문 | Shaw, S.L. (2017). GIS for Transportation, in *The International Encyclopedia of Geography*, 1-9. ¹¹ 논문 | Arentze, T. A., Hofman, F., & Timmermans, H. J. (2001). Deriving rules from activity diary data: A learning algorithm and results of computer experiments. *Journal of Geographical Systems*,3(4), 325-346. doi:10.1007/s101090100069 ¹² 논문 | 임진기, 이광섭, 유소영, 민재홍, (2016). 빅데이터기반 통행자분석시스템 (ABATA) 개발, 대한토목학회학술대회, 2016 (10), 34-35. ¹³ 참고 | 수도권교통본부는 1996년부터 약 5년에 한 번씩 수도권 가구의 약 2 - 3%의 샘플을 추출하여 가구원 전체가 하루 동안 수행한 통행의 상세 정보를 기록한 설문 자료를 수집한다. 이용할 수 있는 최근의 자료는 2010년 조사 자료이다. 개인과 가구의 통행의 특성을 연계하여 분석할 수 있는 현재로서는 최고의 자료이나, 전체의 일부만을 샘플한 점, 특히 자료 내용이 개인 의 기억에 전적으로 의존하는 점 등의 문제를 갖는다. 스마트카드는 수도권 대중교통 이용자가 교통요금 결제 시 제공하는 통행 정보를 일별로 전수 수집한 자료이다. 매일 400만 명 이상, 1000만 건 이상의 통행 기록이라는 엄청난 양의 자료가 아카이브되고 있으나 그 활용은 미미한 실정이다. 주된 이유는, 이 자료가 각 통행의 출발 및 도착 시간과 위치, 이용 교통 수단, 통행요금 등을 분 초 단위까지 정확하게 빠짐 없이 알려주고 있으나, 통행 주체의 특성, 통행의 목적 등 관련 정보를 전혀 제공하지 않기 때문이다. 대중교통 이용의 전수 자료라는 의미, 또 매우 정확하고 빠짐 없는 기록이라는 장점에도 불구하고 실제 통행 이유 등과 관련된 인과적 분석에 대한 이용이 매우 제한적이다. ¹⁴ 논문 | 천승훈, 김현명. (2013). 카 내비게이션 자료의 교통분야 활용방안, 월간교통, 5, 54-59. ¹⁵ 논문 | 이백진, 윤서연, 이춘용, 오성호. (2014). 교통정책 실효성 제고를 위한 활동기반 시뮬레이션 모형개발 및 적용방안, 국토연구원. ¹⁶ 논문 | Lee, W. D., Cho, S., Bellemans, T., Janssens, D., Wets, G., Choi, K., & Joh, C. (2012). Seoul activity-based Model: An Application of Feathers Solutions to Seoul Metropolitan Area. *Procedia Computer Science*, 10, 840-845. doi:10.1016/j.procs.2012.06.109