

# 공공자전거 주행경로 선택모형 분석을 위한 시각화 및 상호작용 설계\*

신동화<sup>1</sup> 조재민<sup>1</sup> 김보형<sup>2</sup> 송현주<sup>3</sup> 서진욱<sup>1</sup>

<sup>1</sup>서울대학교 컴퓨터공학부 <sup>2</sup>한국외국어대학교 바이오메디컬공학부 <sup>3</sup>송실대학교 컴퓨터학부  
dhshin@hcil.snu.ac.kr, jmjo@hcil.snu.ac.kr, bkim@hufs.ac.kr, hsong@ssu.ac.kr, jseo@snu.ac.kr

## Designing Visualizations and Interactions for Public Bicycle Route Choice Modeling

DongHwa Shin<sup>1</sup> Jaemin Jo<sup>1</sup> Bohyoung Kim<sup>2</sup> Hyunjoo Song<sup>3</sup> Jinwook Seo<sup>1</sup>

<sup>1</sup>Department of Computer Science and Engineering, Seoul National University

<sup>2</sup>Division of Biomedical Engineering, Hankuk University of Foreign Studies

<sup>3</sup>School of Computer Science and Engineering, Soongsil University

### 요약

본 연구에서는 교통 공학 분야에서 널리 연구되고 있는 경로 선택모형 분석을 위한 시각화와 상호작용 기법을 제시한다. 특히, 최근 널리 보급되고 있는 공공자전거 이용자의 주행 데이터를 대상으로 연구를 진행하였다. 효과적인 경로 선택모형 분석을 위해 첫째로, 자전거 주행 데이터를 탐색하고 조건에 따라 필터링할 수 있는 시각화를 제안한다. 둘째로, 모형의 추정 시 설정하는 초매개변수들에 따른 다양한 결과를 한 눈에 파악할 수 있는 개요 시각화를 설계하였다. 마지막으로, 경로의 각 속성들이 갖는 계수의 추정치가 추정되는 데에 영향을 준 경로 선택 데이터들이 무엇인지 추론할 수 있는 방법과 시각화를 제안한다. 우리가 제안하는 시각화 및 상호작용을 통해 경로선택 모형을 분석하는 교통 공학 연구자들이 기존에는 발견할 수 없었던 통찰을 효과적으로 이끌어낼 수 있을 것으로 기대된다.

### 1. 서론

경로 선택모형 분석[1]이란 보행자 혹은 운전자가 이동할 경로를 선택함에 있어서 경로의 어떠한 속성을 주로 고려하는지, 그리고 얼마나 고려하는지를 통계 모형으로 만들어 이해하고자 하는 분석 방식이다. 예를 들어, 자전거 주행자의 경우, 경사가 심한 경로보다는 완만한 경로를 선호할 것으로 예상되고, 일반적으로 긴 경로보다는 짧은 경로를 선호할 것이 예상된다. 실제 자전거 주행자들의 경로 선택 데이터를 통해 ‘거리’, ‘경사’, 혹은 ‘횡단보도 개수’ 등 경로가 가질 수 있는 속성들에 대한 계수(coefficient)를 추정하고, 통계적 유의성을 따져봄으로써 앞서 서술한 가설들을 검증할 수 있다.

교통 공학 연구자들은 이 분석의 수행에 있어 다음의 어려움이 있다.

- 모형 추정(model estimation)에 앞서, 연구자들은 다양한 측면에서 데이터를 이해하고, 그 과정에서 얻은 통찰을 이용해 원하는 데이터만을 추려내어 분석할 필요가 있다. 그러나 이 과정이 생략되거나, 혹은 GIS 도구를 활용해 데이터의 지리적 분포만을 확인하는 경우가 대부분이다.
- 모형을 추정하기 까지 여러 초매개변수의 설정이 필요한데, 이것들을 어떻게 설정하느냐에 따라 최종 추정치의 결과가 크게 달라질 수 있다[2]. 그러나 다양한 설정에 대한 모형 추정을 효율적으로 수행할 수 있게 해주고, 그에 따른 여러 결과들을 비교할 수 있는 수단이 부재하다.
- 추정 결과에 해당하는 계수들을 표면적으로 해석하는 것을 넘어, 추정에 활용한 데이터 중 과연 어떤 관측 값이 이러한 계수 값으로 추정되는 데에 공헌을 하였는지 알 수 있다

면 더욱 깊이 있는 분석이 가능하다. 그러나 현재는 이런 분석을 수행할 수단이 부재하다.

따라서 우리는 이러한 기존 분석 방식의 한계를 극복하여 더 효과적인 분석을 할 수 있게 해주는 시각화 및 상호작용을 설계하여 이를 제시하고자 한다.

### 2. 경로 선택모형

경로 선택모형은 교통 공학에서 매우 널리 알려진 문제이다. 그 중에서도 본 연구는 공공자전거의 주행경로에 대한 선택모형을 다루므로, 이에 맞춰 그 배경 지식을 간단히 서술하고자 한다.

**[용어 및 배경지식]** 공공자전거 이용자들은 정해진 대여소에서 자전거를 빌리거나 반납하며, 이 대여소는 도시 곳곳에 분포되어 있고 고정되어 있다. 그러므로 단일 주행(trip)은 전체 대여소 중 한 곳에서 시작되어, 다른 한 곳에서 끝난다. 편의상 단일 주행 데이터를 이야기할 때 그 주행 데이터의 출발 대여소를 O(Origin)로, 도착 대여소를 D(Destination)라고 지칭한다. 또한 단일 주행은 특정 O와 D 사이의 경로 중 하나를 선택한 주행경로 선택 관측 값의 하나로 취급할 수 있다.

**[대안경로 생성]** 자전거 이용자가 특정한 O와 D간의 하나의 경로를 선택할 때, 주행 가능한 많은 경로 중 대표 경로 몇 가지를 선택 가능 경로들을 고정하는 것을 대안경로 생성이라고 한다. 본 연구에서는 각 OD(특정 O와 D 한 쌍) 별로 자전거 이용자들이 실제 주행한 경로(observed route)들을 모아 k-메도이드(k-medoids) 군집화를 사용하여 k개의 대표 경로를 추출한다. 여기서 설정할 초매개변수는 경로 간 유사도 측정에 쓰이게 될 경로 속성 집합(예시로 {‘거리’, ‘횡단보도 개수’} 등), 그리고 대표 경로의 개수에 해당하는 k값이다.

**[모형의 추정]** 이 연구에서 사용하는 모형은 path-size logit 모형[3]이다. 추정할 경로 속성들을 정하면, 그 속성들에 대한

\* 이 논문은 2020년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No. NRF-2019R1A2C2089062).

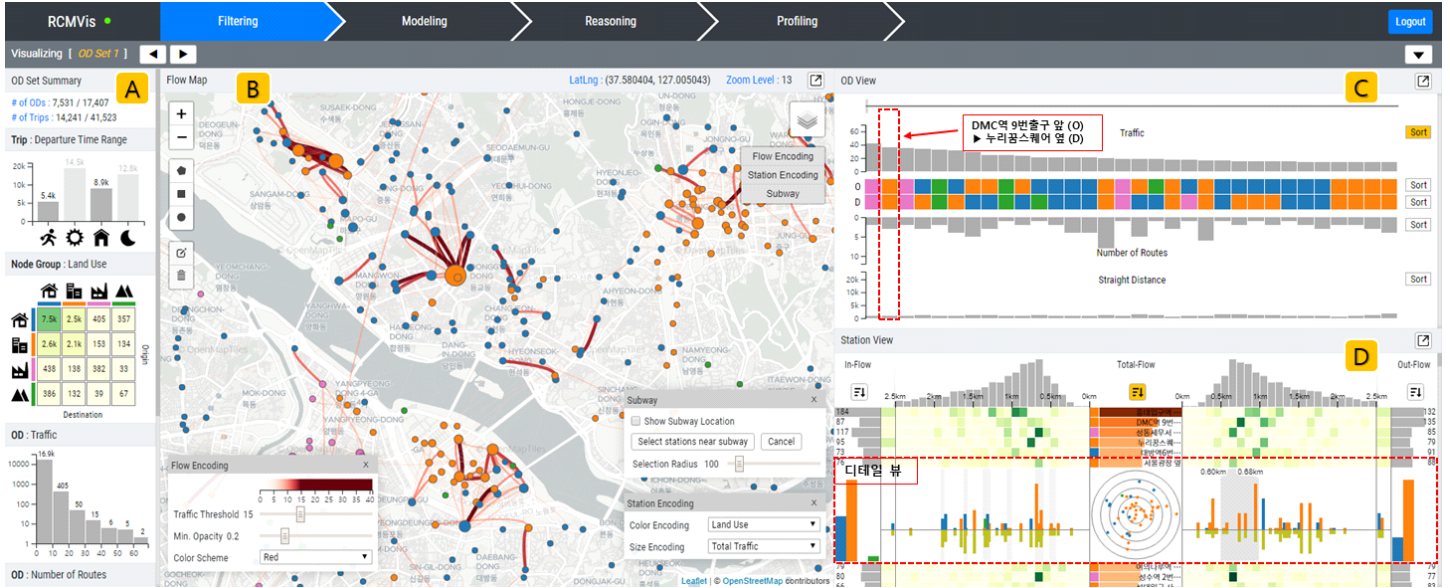


그림 1. 필터링 과정 인터페이스이다. A는 사용자가 데이터에 다양한 필터를 적용할 수 있는 필터 뷰이다. B는 데이터의 지리적 분포를 확인할 수 있는 흐름 지도이다. C는 OD 단위의 각 속성 별로 데이터를 탐색할 수 있는 OD 뷰이다. D는 자전거 대여소 단위로 얼마나 많은 통행량이 있는지, 그것이 어떤 OD인지 확인할 수 있는 대여소 뷰이다.

각각의 계수가 추정이 되고, 또한 p-value를 통해 통계적 유의성 또한 조사할 수 있다. 모델의 적합도(goodness-of-fit) 측정을 위해 rho-bar-squared라는 값을 확인할 수 있으며, 이 값은 1 이하의 범위를 갖는다.

### 3. 시각화 및 상호작용 기법

경로 선택모형의 효과적인 분석을 위해 우리는 아래와 같이 분석을 개념에 따라 3단계로 나누어 수행할 수 있게 하였다.

1. **필터링 과정** 데이터를 탐색하고 원하는 제약 조건을 걸어 분석에 활용할 수 있는 데이터를 도출해낸다.
2. **모델링 과정** 선택된 데이터에 다양한 초매개변수를 설정하여 대안경로 생성 ▶ 모델 추정까지 수행하며, 다양한 결과들을 확인, 비교할 수 있다.
3. **결과 추론 과정** 모델 추정 결과 중 특정 추정치가 어떤 경로 선택 관측 값에 영향을 받았는지 추론하여 한 단계 더 깊이 있는 분석을 수행할 수 있다.

**[필터링 과정]** 그림 1은 필터링 과정에 해당하는 사용자 인터페이스이다. 주요 시각화는 네 가지로 아래에서 서술한다.

**필터 뷰:** 사용자는 그림 1-A의 필터 뷰(view)를 이용해 다양한 조건으로 데이터에 필터를 적용할 수 있다. 주된 조건으로는 주행 출발 시간, O 혹은 D의 토지 이용 유형(land use), OD의 통행량(traffic), OD의 경로 가짓수(number of routes), 그리고 OD의 거리(distance)가 있다. 이 중에서 특히, 토지 이용 유형의 경우 인터페이스 전반에 걸쳐서 고유한 색깔로 나타내고 있다. 이는 그림 1-A 중단의 행렬(matrix) 시각화에서 확인할 수 있다. 사용자는 각 막대그래프 혹은 행렬 시각화의 막대나, 행, 열, 혹은 셀(cell)을 선택함으로써 필터링이 가능하다. 이렇게 전체, 혹은 필터링 된 데이터는 흐름 지도(flow map), OD 뷰(OD view), 그리고 대여소 뷰(station view)를 통해 표현된다.

**흐름 지도:** 자전거 주행자들의 이동을 경로 수준이 아닌, 대여소 간의 흐름으로 모아서(aggregation) 나타낸 시각화[4]이다(그림 1-B). 주행 데이터의 지리적 분포를 매우 직관적으로 알 수 있다는 강점이 있다.

**OD 뷰:** 특정 O와 D를 하나의 분석 단위로 보고, 데이터 상

의 OD들의 정보를 시각화한다(그림 1-C). 이 뷰는 “공공자전거가 가장 많이 이용되는 코스(OD)는 어디일까?”와 같은 질문에 답을 할 수 있다. 시각화 상의 열(column)이 하나의 OD를 나타내며, 위에서부터 순서대로 통행량, O와 D의 토지 이용 유형, 경로 가짓수, 그리고 거리를 나타내는 차트들로 구성되어 있다. 흐름 지도만을 통해서서는 정확하게 파악하기 힘든 특정 OD의 여러 속성들을 기준으로 정렬하여 볼 수 있다는 강점이 있다.

**대여소 뷰:** 특정 자전거 대여소를 기준으로한 통행량이나 움직임을 분석하기 위한 시각화이다(그림 1-D). 각 행(row)은 대여소 한 군데에 대한 정보를 나타내며, 행의 중앙에는 대여소의 전체 통행량과 토지 이용 정보, 그리고 이름이 있다. 각 행의 좌우로는 각각 해당 대여소로 반납을 하러 오는 움직임들과, 그 대여소에서 대여해서 다른 곳으로 이동하는 움직임을 거리에 따라 비닝(bin)하여 나타내고 있다. 따라서 대여소 뷰를 보면 특정 대여소가 주로 어느 정도 거리에 있는 곳으로 가는 혹은 오는 통행량이 많은지 알 수 있는 것이다.

더욱 자세한 정보는 각 행이 선택되었을 때, 아래에 디테일 뷰(detail view)에서 제공된다. 디테일 뷰는 통행량 뿐 아니라, 정확히 어느 대여소와의 움직임이 얼마나 있는 것인지 막대를 통해 나타낸다. 이 막대 하나는 하나의 OD를 나타내는 것으로서, OD 뷰의 열 하나와 대응될 수 있다. 막대의 상단은 통행량을, 하단의 노란 부분은 경로 가짓수를 나타낸다. 이 움직임들에 대한 거리 뿐 아니라 방향을 알기 위해 중단에 레이더 모양의 뷰를 두어서, 사용자가 해당하는 막대에 마우스를 올리면 레이더 상에서 해당 하는 점이 강조되도록 설계하였다.

**[모델링 과정]** 그림 2는 모델링 개요(overview)로서, 필터링 과정에서 사용자가 원하는 필터가 적용된 데이터로부터 대안경로를 생성한 결과와, 그에 대한 모델 추정 결과까지 한눈에 확인할 수 있는 시각화이다. 그림 2-A는 사용자가 필터링 과정에서 어떤 필터들을 적용했는지 알 수 있게 해주는 글리프(glyph) 형태의 시각화이다. 그림 2-B는 대상 OD 데이터들로부터 k를 2부터 6까지 바꿔가며 대안경로 생성을 한 결과들을 나타내고 있다. 따라서 서로 다른 글리프 5개가 만들어져 연결된 모습이다.

서로 다른 대안경로 결과에 대해 각각 모형 추정을 시행했을 때 그림 2-C와 같이 그 결과가 표 형태 나타난다. 각 행은 모

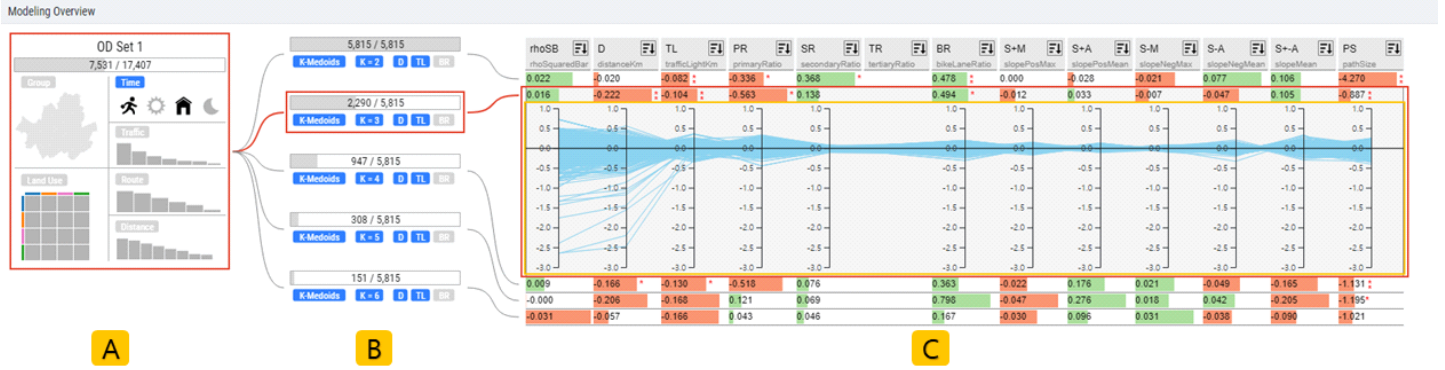


그림 2. 모델링 과정 인터페이스이다. A에서는 필터링 과정에서 적용된 필터를 확인 가능하며, 서로 다른 설정을 통해 다양한 대안경로 생성 결과를 B에서 글리프로 확인 가능하다. C에서는 B의 결과들에 대한 모델 추정 결과들을 표 형태로 나타낸다.

델 추정 결과에 해당하며, 이 또한 해당되는 대안경로 결과 글리프와 연결되어 있어, 데이터 - 결과 관계를 쉽게 파악할 수 있다. 각 열은 추정의 대상인 경로 속성이다. 계수의 부호는 해석에 매우 중요한 역할을 하는데, 양수의 경우 초록, 음수의 경우 빨간 색으로 표현하였다. 또한 계수의 크기 비교를 위해 각 셀들에 길이가 다른 막대를 배치하였다. 통계적 유의성 표현을 위해 셀마다 빨간 별(asterisk)을 배치하였는데 한 개는 ( $p < .05$ ), 두 개는 ( $p < .01$ )를 의미한다.

**[결과 추론 과정]** 그림 2-C의 결과들 중에서, 위에서 두 번째 결과는 '거리'의 계수가 -0.222로 상대적으로 그 절대 값이 크고, 통계적으로 유의했다( $p < .01$ ). 이는 공공자전거 이용자들이 여러 경로들 중 거리가 짧은 경로를 비교적 강하게 선호함을 의미한다. 이 때, 이러한 계수가 추정되는 데에 공헌을 한 데이터가 과연 무엇인지 살펴봄으로써 표면적 해석이 아닌 데이터 수준에서의 깊이 있는 해석을 필요로 할 수 있다. 그럴 때 해당하는 결과 행을 표에서 선택하면 그림 2-C와 같이 그 결과 행 아래 평행좌표계(parallel coordinates)가 나타난다. 각 축은 행과 마찬가지로 경로 속성을 나타내지만 그 값은 해당 계수에 대한 공헌 점수를 의미한다. 따라서 평행좌표계의 '거리'축에서 상위에 있는 선들만을 살펴본다면, 앞서 계수가 -0.222로 음의 방향으로 상당히 크게 나온 원인이 되는 데이터를 살펴보는 것이나 마찬가지인 셈이 된다.

그림 3과 같이 평행좌표계의 각 축 별로 원하는 값을 선택할 수 있도록 브러싱(brushing)을 제공한다. 그에 따라 앞서 필터링 과정에서 소개된 OD 뷰가 다시 활용되며, 각 OD 별로 속성들에 대한 공헌 점수를 색깔로 나타낸다. 그림 3에서는 '거리'에 대해 내림차순으로 정렬 한 뒤 가장 공헌도가 높은 OD를 선택한 모습이다. 그러면 사용자가 해당하는 OD의 지도 상 실제 경로를 확인할 수 있는 경로 지도가 나타나게 되는데, 다

양한 경로가 서로 다른 색으로 표시된다. 이 중 O와 D 사이 다른 긴 경로가 아닌 보라색의 짧은 경로를 선택한 관측 값들이 많이 존재하며, 바로 이러한 선택들이 계수 추정과정에서 거리를 상대적으로 크기가 큰 음의 계수로 나오는 데에 공헌했음을 확인할 수 있다.

#### 4. 결론 및 향후 연구

본 연구를 통해 교통 공학 분야의 유명한 문제인 경로 선택 모형 분석을 효과적으로 수행할 수 있게끔 세 단계로 나누어 그에 따른 데이터 시각화 및 상호작용의 설계를 제시하였다. 향후에는 시각화의 설계에 대한 정성적, 정량적인 평가를 수행하여 더욱 그 설계를 개선할 수 있을 것으로 생각된다.

#### 참고 문헌

- [1] Prato, C. G. Route choice modeling: Past, present, and future research directions, *Journal of Choice Modeling*, 2(1), 65-100, 2009.
- [2] Pagliara, F., Timmermans, H. Choice set generation in spatial contexts: a review, *Transportation Letters*, 1(3), 181-196, 2009.
- [3] Bernardi, S., La Paix-Puello, L., Geurs, K. Modeling route choice of Dutch cyclists using smartphone data, *Journal of Transport and Land Use*, 11(1), 883-900, 2018.
- [4] Wood, J., Slingsby, A., Dykes, J. Visualizing the Dynamics of London's Bicycle-Hire Scheme, *Cartographica*, 46(4), 239-251, 2011.

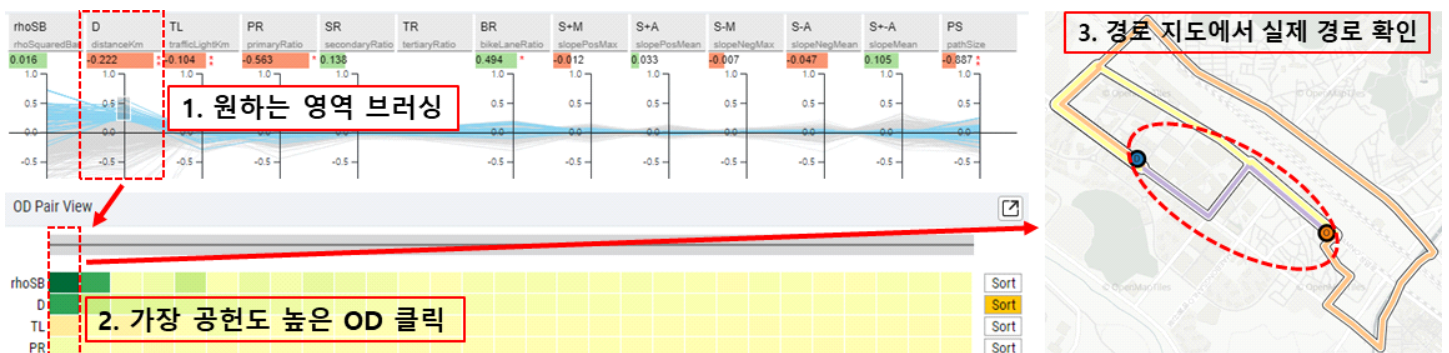


그림 3. 결과 추론 과정을 요약하고 있다. 먼저 평행좌표계에서 원하는 축의 특정 부분을 브러싱하면, 그에 맞춰 OD 뷰가 필터링되고, 그 중 원하는 열을 클릭하면 해당 OD의 경로 지도가 나타나서 3번과 같이 실제 경로를 지도상에서 확인할 수 있다.