

자연어처리 기반 인지적 도달거리 분석: 서울시 아파트 실거주자 리뷰를 중심으로

조성연*, 박정진*, 정진희**, 이경주*
*국립한국교통대학교 도시·교통공학전공
**국립한국교통대학교 도시·교통공학과
e-mail:csy10034@naver.com

Analyzing Cognitive Threshold Distances of High-order Facilities in Seoul Apartments: Focusing on LLM-based Sentiment Analysis and Piecewise Regression

Sung-yeon Cho*, Jung-jin Park*, Jin-hee Jung**, Gyoung-ju Lee*
*Dept. of Urban & Transportation Engineering, Korea National University of Transportation
**Dept. of Urban & Transportation Engineering, Korea National University of Transportation

요약

이 연구는 서울시 아파트 거주자의 직방 리뷰 데이터를 기반으로 도시 시설에 대한 인지적 임계 거리(cognitive threshold distance)를 시설 위계별로 규명하는 것을 목적으로 한다. 직방 리뷰 데이터를 바탕으로 KoELECTRA 기반 속성 기반 감성 분석(ABSA)을 통해 시설 유형별 긍정 언급을 추출하고, 단지와 시설 간 최단 거리에 분할 선형 회귀를 적용하여 변곡점을 인지적 임계 거리로 해석하였다. 분석 결과, 공원·공공도서관 등 일상 시설은 단지 인접 수준(약 0.1km 이내), 1차 병원·복합쇼핑몰 등은 도보권(약 0.4~0.6km), 2·3차 병원·백화점·전통시장 등 고차위 시설은 광역 생활권(약 1.3km)의 임계 거리를 갖는 것으로 나타났다. 또한 다수 시설에서 V자형 비단조 패턴이 관측되어, 거주자의 시설 인지가 접근편익과 외부불경제의 상쇄 메커니즘에 의해 형성됨을 확인하였다. 이 연구는 전통적 중심지 이론의 도달거리 개념을 거주자의 인지적 차원으로 확장하고, 도시 계획 및 주거 입지 분석에 활용 가능한 실증적 지표를 제시함으로써 대도시 공간 구조 연구에 기여할 것으로 기대된다.

1. 서론

1.1 연구의 배경 및 목적

전통적인 중심지 이론(Central Place Theory)에 따르면, 기업의 상권 범위는 재화의 가격(생산가격과 교통비의 합)에 의해 결정된다. 소비자와 기업 간의 거리가 멀어질수록 재화 구입에 소요되는 교통비가 증가하므로 소비자의 실질구매비용은 증가하게 되며, 이로 인해 시장에서의 재화요량은 점차 감소한다(구동회, 2017). 일정 수준의 거리에 도달하면 수요가 더 이상 발생하지 않는 지점이 형성되는데, Christaller(1933)는 이 지점을 재화의 도달거리(range of good)로 정의하였으며, 이를 "주민들이 중심 재를 구입하기 위하여 기꺼이 이동하고자 하는 가장 먼 거리"로 규정하였다(구동회, 2017). 재화의 도달거리를 360°로 회전시켜 형성되는 영역이 곧 기업의 상권이 되며, 따라서 도달거리는 소비자에게 재화를 제공할 수 있는 최대 거리라 할 수 있다(이희연, 2011).

그러나 이러한 고전적 중심지 이론은 대도시 환경에서 그 설명력이 제한된다는 한계가 지적되어 왔다. 대도시는 기능 집적과 인구 밀도 증가로 인해 중심지 기능 간 공간 경쟁이 발생하고, 이로

인해 이론이 가정하는 이격(spacing) 효과가 충분히 발휘되지 못하기 때문이다(허우궁, 2004). 특히 서울과 같은 대도시는 대중 교통 체계의 고도화로 교통비 부담의 영향이 상대적으로 약화되어, 단순한 거리-교통비 함수만으로는 거주자의 시설 이용 행태를 충분히 설명하기 어렵다.

이러한 한계를 보완하기 위해 도시 연구에서는 인지적 거리(cognitive distance)의 개념이 주목받아 왔다. 인지적 거리는 객관적 유클리드 거리와 달리, 거주자가 환경을 인식하고 해석하는 과정에서 형성되는 주관적 거리로, 동일한 물리적 거리라도 도시 형태와 환경 요소에 따라 체계적인 왜곡이 발생하는 것으로 알려져 있다(Manley et al., 2021). 인지적 거리는 거주자의 공간적 의사결정과 행태에 직접적인 영향을 미치며, 이는 곧 거주자가 시설에 부여하는 인식과 평가에 반영된다(Lloyd, 1989; Manley et al., 2021).

이러한 맥락에서 거주자가 자신의 주거 환경에 대해 직접 작성한 리뷰 데이터는 단순한 지리적 좌표 정보를 넘어, 거주자의 주관적 인식과 경험을 반영한 자료로서 가치를 가진다. 최근 자연어 처리(Natural Language Processing) 기술의 발전, 그중에서도 속성 기반 감성 분석(Asspect-Based Sentiment Analysis, ABSA) 기법은 비정형 텍스트 내에서 특정 속성에 대한 감성을 개별적으로 추출할 수 있게 하여, 리뷰 데이터로부터 정량적 지

표를 도출하는 것을 가능하게 하였다(장호준 외, 2025). 이러한 분석 기법은 기존의 통계적 지표만으로는 포착하기 어려웠던 거주자의 주관적 시설 인식을 분석할 수 있는 방법론적 기반을 제공한다.

이러한 배경을 바탕으로, 이 연구의 목적은 인지적 거리를 재화의 도달거리에 대한 새로운 측정 기준으로 제시하는 것이다. 구체적으로, 서울시 아파트 실거주자 리뷰에서 특정 시설이 긍정적으로 언급되는 빈도와 단지로부터의 거리 간의 관계를 분석함으로써, 거주자가 해당 시설을 자신의 생활권 내 시설로 인식하는 임계 거리를 도출하고자 한다. 이러한 연구를 통해 전통적 도달거리 개념을 거주자 인지의 차원으로 확장하고, 거주자의 시설 인식 행태에 기반한 실증적 도시 분석에 기여하고자 한다.

1.2 연구의 범위

이 연구는 부동산 플랫폼 '직방'의 아파트 실거주자 리뷰 데이터를 활용하였다. 공간적 범위는 서울특별시 전역의 아파트 단지로 설정하였으며, 시간적 범위는 주거 환경 변화의 최신성을 반영하기 위해 2021년부터 2026년까지의 5개년 데이터로 한정하였다. 리뷰 데이터는 Python 기반의 웹 크롤링을 통해 수집하였으며, 직방 리뷰는 거주자 1인이 '추천 점수', '교통 여건', '주변 환경', '단지 관리', '거주 환경' 등 5가지 세부 항목을 한 세트로 작성하는 구조적 특징을 지닌다. 본 연구에서는 데이터의 신뢰성과 객관성을 확보하기 위해 상위 5종의 평가 항목이 모두 포함된 리뷰만을 선별하였으며, 단지별로 유효 리뷰가 최소 3인 이상 누락된 경우만을 최종 분석 대상으로 확정하여 데이터의 대표성을 제고하였다. 시설 데이터는 국가공간정보포털의 공공 인허가 데이터에서 시설 유형별 세부 분류를 활용하여, 각 시설 유형별 위계와 인지적 거리를 분석하였다.

2. 분석 방법

2.1 데이터 수집 및 전처리

이 연구는 분석의 정확도를 높이기 위해 [표 1]과 같은 단계적 전처리 과정을 수행하였다. 먼저, 수집된 아파트 리뷰 데이터에 대해 바른AI를 활용한 맞춤법 검사를 진행하였다. 바른AI는 딥러닝 기반의 한국어 형태소 분석 및 맞춤법 교정 모델로, 오타자를 정밀하게 교정하여 데이터의 신뢰성을 제고하는 역할을 수행한다. 다음으로, 맞춤법이 교정된 문장을 대규모 언어 모델(LLM)에 입력하여 문맥에 부합하지 않는 단어를 수정하고, 불용어를 제거하며 의미가 불분명한 단어들을 전반적인 문맥에 맞게 다듬는 작업을 수행하였다.

Step	Data
원시 데이터	특히나 내돈주고 는 절대 사고싶지않음 바카바카 너무 많다.
바른 AI	특히나 내 돈주고 는 절대 사고 싶지 않음 발가 너무 많다.
대규모 언어 모델 (LLM)	특히나 내 돈 주고는 절대 사고 싶지 않음 바카바카 너무 많다.

[표 1] 텍스트 데이터 전처리 절차

2.2 속성 기반 감성 분석을 통한 시설 언급 추출

전처리된 데이터를 바탕으로 KoELECTRA 기반 속성 기반 감성 분석(ABSA)을 수행하였다. KoELECTRA는 한국어 문맥 이해에 특화된 사전학습 언어모델이다(Park, 2020). 속성 기반 감성 분석은 문장 전체의 감성을 일괄 판단하는 일반 감성 분석과 달리, 문장 내 특정 속성(aspect)—이 연구에서는 '병원', '백화점', '학교' 등 시설명—을 추출하고 해당 속성에 대한 감성 극성(긍정/부정)을 개별적으로 판별하는 기법이다(장호준 외, 2025). 이 연구에서는 ABSA를 다음과 같이 두 단계로 활용하였다. 첫째, 속성 추출(aspect extraction) 단계에서 리뷰 문장 중 특정 시설 유형이 언급된 문장을 식별하였다. 둘째, 감성 분류(sentiment classification) 단계에서 해당 시설이 언급된 문맥이 긍정적인지 여부를 판별하여, 긍정 맥락에서 언급된 시설만을 분석 대상으로 선별하였다. 이를 통해 단순 키워드 매칭의 한계—예컨대 "병원이 너무 멀어서 불편하다"와 같은 부정적 언급—를 배제하고, 거주자가 실제로 자신의 생활권 시설로 인식한 경우만을 추출하였다.

2.3 분할 선형 회귀를 통한 인지적 임계 거리 도출

각 아파트 단지와 시설 유형별 가장 가까운 시설 사이의 최단 직선 거리를 산출한 후, 분할 선형 회귀(Piecewise Linear Regression)를 적용하여 인지적 임계 거리를 도출하였다. 분석 단위는 '아파트 단지 × 시설 유형'으로 설정하였으며, 독립변수는 단지에서 가장 가까운 해당 시설까지의 최단 직선 거리(km)이고, 종속변수는 KoELECTRA 기반 ABSA에서 해당 시설이 긍정 맥락으로 추출된 리뷰들의 단지별 평균 감성 점수(1~5점)이다. 즉 본 연구는 거리에 따라 거주자가 해당 시설에 부여하는 인지 만족도가 어떻게 변화하는지를 모형화하였다. 분할 회귀(Piecewise Regression)는 두 변수 간 관계의 기울기가 급격히 변화하는 지점, 즉 변곡점(breakpoint)을 데이터로부터 추정하는 통계 기법이다(Muggeo, 2003). 본 연구에서는 두 개의 선분(2-segment)으로 구성된 분할 회귀 모형을 적합하여, 거리 증가에 따른 시설 인지 만족도의 기울기가 가장 크게 변화하는 지점을 변곡점으로 추정하였다. 본 연구는 이 변곡점을 인지적 임계 거리(cognitive threshold distance)로 정의하고, 거주자의 시설 인지 만족도가 거리에 의해 분기되는 거리의 한계로 해석하였다. 기존의 도달거리 개념이 '수요가 발생하는 최대 거리'를 의미한다면, 본 연구의 인지적 임계 거리는 '거주자가 해당 시설을 자신의 생활권 시설로 인지하는 만족도 변환점에 해당한다. 이는

객관적 거리-수요 관계가 아닌, 주관적 인식-만족도 관계를 통해 도시 시설의 도달 범위를 재해석한 것이다.

3. 분석 결과

3.1 시설 유형별 인지적 임계 거리 도출

각 시설 유형에 대해 단지 단위의 최단 거리를 독립변수로, 긍정 리뷰 내 언급 발생 여부에 기반한 시설 만족도를 종속변수로 설정한 분할 회귀 모형을 적합하였다. 변곡점(breakpoint)을 인지적 임계 거리로 해석한 결과는 [표 2]와 같다.

위계	시설	표본수	임계거리 (km)	기울기1	기울기2
일상시설 (단지인접)	공원	352	0.059	-12.346	0.095
	공공도서관	75	0.107	-6.204	0.506
	어린이집	61	0.593	-1.321	0.094
도보권 시설	1차병원 (의원급)	34	0.416	-2.621	0.478
	영화관	45	0.649	-1.385	0.468
	복합쇼핑몰	97	0.403	-1.620	0.144
광역생활권 시설	2차병원 (종합병원)	185	1.387	-0.507	0.282
	3차병원 (상급종합병원)	28	1.315	-0.748	0.108
	백화점	222	1.288	0.203	-0.035
	전통시장	209	1.281	-0.269	0.180
	관광숙박업	45	2.262	-0.207	0.261

[표 2] 시설 유형별 인지적 임계 거리 도출 결과

주 1) 기울기1은 임계 거리 이전 구간(근거리)의, 기울기2는 임계 거리 이후 구간(원거리)의 회귀 기울기이며, 음(-)은 거리 증가에 따라 만족도가 감소함을, 양(+)은 증가함을 의미한다.

주 2) 분할 회귀의 적합도가 낮거나(표본 25 미만, 기울기의 절댓값이 비정상적으로 큰 경우), 거리 변동이 거의 없는 시설(슈퍼마켓, 편의점, 카페, 미용실 등 도보권 내 균질 분포 시설)은 결과 해석의 신뢰도가 제한되어 본 표에서 제외하였다.

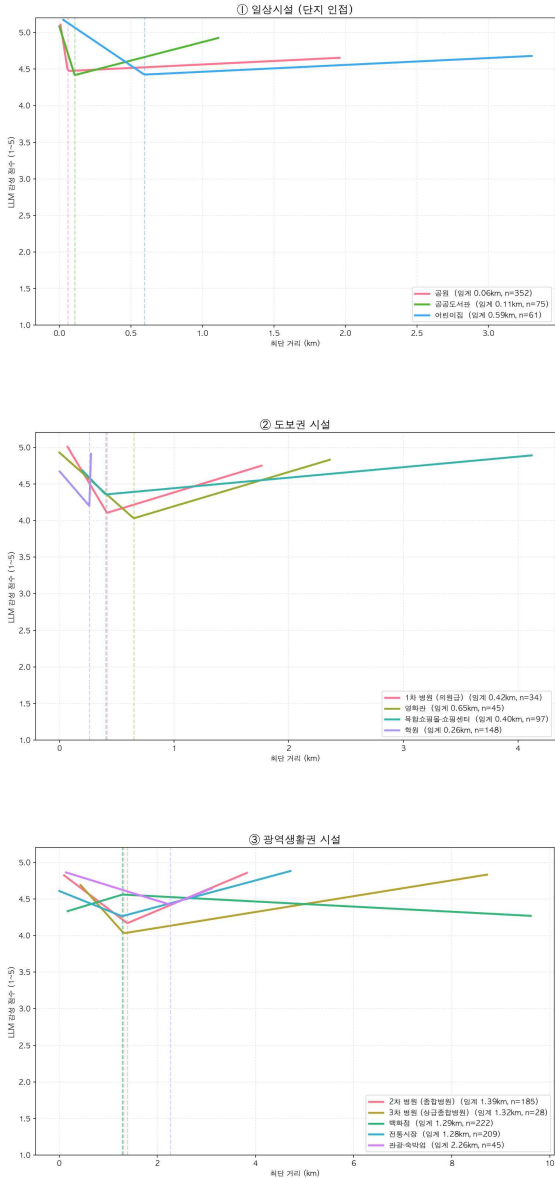
분석 결과, 시설의 위계와 기능적 성격에 따라 인지적 임계 거리에 뚜렷한 차이가 나타났다(자세한 분포는 [그림 1]~[그림 3] 참조). 첫째, 보건의료 시설은 위계에 따른 임계 거리의 차이가 분명하게 확인되었다. 1차 병원은 약 0.42km, 2차 병원은 약 1.39km, 3차 병원은 약 1.32km로 도출되어, 1차 병원은 도보 5분 권역의 강한 근접 선호를, 종합병원 이상은 도보·교통수단을 결합한 광역적 인지 범위를 보였다. 이는 거주자가 시설의 기능적 위계에 따라 수용 가능한 거리의 한계를 차등적으로 설정함을 시사한다. 둘째, 일상 생활 영역과 밀접한 시설은 매우 짧은 거리에서 변곡점이 형성되었다. 공원(0.06km)과 공공도서관(0.11km) 모두 단지에 직접 면한 초근접 입지에서 거주자가 시설을 강하게 자신의 생활권 시설로 인지하여 높은 만족도가 형성되었으며, 임계 거리에 도달하면서 인지 강도가 급격히 약화되어 만족도가 최저점에 수렴하는 패턴이 관측되었다. 이는 일상적·반

복적 이용 시설일수록 거주자의 인지가 시설과의 직접 인접 여부에 의해 민감하게 분기됨을 시사한다. 셋째, 교육·보육 및 여가 시설은 도보권 내 임계 거리를 갖는 것으로 분석되었다. 어린이집(0.59km), 영화관(0.65km), 복합쇼핑몰(0.40km)이 모두 도보 10~15분 이내로 도출되어, 일상적 통행을 전제로 하면서도 매일 이용하지는 않는 준일상적 시설의 성격을 반영하였다. 넷째, 고차위 상업 시설인 백화점(1.29km)과 전통시장(1.28km)은 1km대의 비교적 넓은 임계 거리를 보였다. 특히 백화점은 첫 번째 구간 기울기가 양(+0.203)의 부호를 보였는데, 이는 단지가 백화점에 지나치게 인접한 경우 야간 영업·차량 출입·교통 혼잡 등에 의한 부정적 외부효과가 접근편익을 상회함을 시사하며 (Rosen, 1974), 고차위 상업 시설의 입지가 단순 거리 함수가 아닌 적정 이격 거리(optimal spacing)를 갖는 비선형 관계임을 보여준다. 다섯째, 관광·숙박업의 임계 거리는 약 2.26km로 가장 광범위하게 도출되어, 비정기적·선택적 이용 대상에 대해 거주자가 보다 관대한 거리 기준을 적용함을 보여준다. 요컨대, 인지적 임계 거리는 시설의 일상성과 위계에 따라 (1) 단지 인접 수준의 매우 짧은 임계 거리(공원, 공공도서관 등 일상 시설), (2) 도보권 임계 거리(보육·교육·여가 시설, 1차 병원), (3) 광역 생활권 임계 거리(2~3차 병원, 백화점, 전통시장, 관광·숙박업)의 세 층위로 분화된다. 한편 1~3차 병원, 공공도서관, 어린이집, 영화관, 복합쇼핑몰, 전통시장, 관광·숙박업 등 대부분의 시설에서 첫 구간 기울기 음(-)·둘째 구간 기울기 양(+)의 V자형 비단조 패턴이 관측되었다. 이는 거주자가 단지에 인접한 시설일수록 이를 자신의 생활권 시설로 강하게 인지하여 높은 만족도를 부여하는 반면, 임계 거리에 도달하면서 인지 강도가 약화되어 만족도가 최저점에 수렴하고, 임계 거리 너머에서는 해당 시설을 더 이상 자신의 생활권 시설로 인식하지 않게 되어 만족도가 거주환경 일반 수준으로 회복되는 인지적 거리 분기 메커니즘으로 해석할 수 있다. 즉 V자형 패턴의 변곡점은 거주자가 해당 시설을 자신의 생활권 시설로 인지하는 거리의 한계, 즉 인지적 임계 거리에 해당한다. 반면 백화점은 첫 구간 기울기가 양(+0.203), 둘째 구간이 음(-0.035)으로 도출되어 다른 시설과 구별되는 역V자 패턴을 보였는데, 이는 백화점에 매우 인접한 단지에서 부정적 외부효과(소음·교통 혼잡·야간 영업·차량 출입 등)의 강도가 다른 시설보다 두드러져 임계 거리(약 1.29km)에 도달할 때까지는 거리 증가가 만족도 회복으로 이어지지만, 그 이후로는 접근편익의 점진적 약화가 우세해지는 외부효과·접근편익 균형 메커니즘으로 해석된다(Rosen, 1974). 두 패턴 모두 거리-만족도 관계가 단조 감소형이 아닌 변곡점을 갖는 비선형 구조임을 공통적으로 보여 주며, 본 연구의 인지적 임계 거리는 '거주자가 해당 시설을 자신의 생활권 시설로 인지하는 거리의 한계로 재정의될 수 있다. 이는 고전적 중심지 이론이 가정하는 단조감소형 도달거리 함수와 구별되는 대도시 환경의 비선형 인지 구조를 보여준다.

생활권으로 인지하는 강도가 거리에 따라 분기되는 인지적 거리 분기 메커니즘으로, 역V자형은 부정적 외부효과와 접근편익이 균형을 이루는 외부효과-접근편익 균형 메커니즘으로 각각 형성됨을 확인하였다(Rosen, 1974). 본 연구의 시사점은 다음과 같다. 첫째, 전통적 중심지 이론의 도달 거리 개념을 거주자 인지의 차원으로 확장하여, 대도시 환경에서의 새로운 측정 기준을 제시하였다. 둘째, 시설 유형별 차등 임계 거리는 향후 도시 계획·생활 SOC 배치·주거 입지 분석에서 활용 가능한 실증적 지표가 될 수 있다. 셋째, 백화점에서 관측된 역V자 패턴은 고차위 상업 시설의 입지 결정에서 단순 근접성이 아닌 적정 완충 거리(buffer distance)의 설정이 필요함을 시사한다. 다만 본 연구는 시설명 키워드 매칭 기반 ABSA의 한계, 일부 시설 유형의 표본 부족, 직선 거리 기반 측정의 한계를 지니며, 향후 문장 단위 ABSA·네트워크 거리 기반 분석(Manley et al., 2021)을 결합한 후속 연구가 요구된다.

참고문헌

- [1] 구동희, "크리스탈러의 중심지이론에서의 Range 개념", 국토지리학회지, vol. 51, no. 4, pp. 391-406, 2017.
- [2] 이희연, "경제지리학(제3판)", 법문사, 2011.
- [3] 장호준·전다원·이희상, "사전학습 언어모델을 사용한 다국어 속성기반 감성분석: 호텔 리뷰를 중심으로", 대한산업공학회지, vol. 51, no. 1, pp. 36-47, 2025.
- [4] 허우궁, "교통의 지리", 한울아카데미, 2004.
- [5] Manley, E., Filomena, G., & Mavros, P., "A spatial model of cognitive distance in cities," International Journal of Geographical Information Science, vol. 35, no. 11, pp. 2316-2338, 2021.
- [6] Christaller, W., "Die zentralen Orte in Süddeutschland," Jena: Gustav Fischer, 1933.
- [7] Lloyd, R., "The estimation of distance and direction from cognitive maps," American Cartographer, vol. 16, no. 2, pp. 109-122, 1989.
- [8] Muggeo, V. M. R., "Estimating regression models with unknown break-points," Statistics in Medicine, vol. 22, no. 19, pp. 3055-3071, 2003.
- [9] Park, J., "KoELECTRA: Pretrained ELECTRA Model for Korean," GitHub, <https://github.com/monologg/KoELECTRA>, 2020.
- [10] Rosen, S., "Hedonic prices and implicit markets: Product differentiation in pure competition," Journal of Political Economy, vol. 82, no. 1, pp. 34-55, 1974.



4. 결론

본 연구는 서울시 아파트 실거주자 리뷰 데이터를 기반으로, 거주자가 자신의 생활권 시설로 인식하는 인지적 임계 거리 (cognitive threshold distance)를 시설 유형별로 도출하는 것을 목적으로 하였다. 이를 위해 KoELECTRA 기반 ABSA로 시설 유형별 긍정 언급을 추출하고, 단지과 시설 간 최단 거리에 분할 선형 회귀를 적용하여 변곡점을 인지적 임계 거리로 해석하였다. 분석 결과, 시설의 위계와 일상성에 따라 인지적 임계 거리는 ① 단지 인접 수준(공원·공공도서관, 100m 이내), ② 도보권(1차 병원·복합쇼핑몰, 0.4~0.6km), ③ 광역생활권(2~3차 병원·백화점·전통시장, 1.3km 내외)의 세 층위로 뚜렷하게 분화되었다. 또한 분석 대상 시설 대부분에서 V자형(다수 시설) 또는 역V자형(백화점) 비단조 패턴이 관측되어, V자형은 거주자가 시설을 자신의