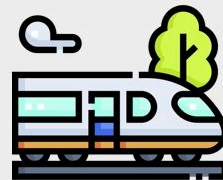


서울시 역세권의 토지 유형별 TOD 균형 개발 방안 제시



내학점구해조

산업공학과 2020147002 박유정
산업공학과 2020147030 홍승은
산업공학과 2020147035 정주원
산업공학과 2020147039 유준형



지도교수 이 원 상
조교 전 은 지



Contents

Introduction

- TOD의 개념
- 연구 배경
- RQ

Literature Review

- 선행연구 고찰
- 프로젝트의 차별성

Data Description

- 데이터 출처
- 데이터 특성
- 데이터 변수 제시

Methodology

- Preprocessing
- SNA
- Clustering
- PCA
- Multi Regression

Conclusion

- 종합분석
- 방안 제시
- 한계점

References

- 참고문헌
- 데이터출처



Abstract

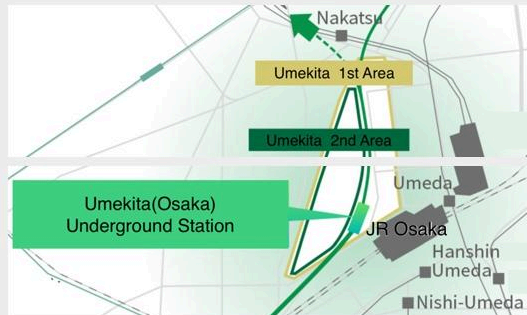
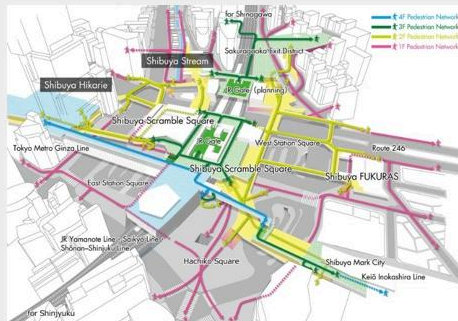
본 프로젝트는 서울시 역세권의 TOD 균형 개발 방안을 제시하는 것을 목적으로 한다.

최근, 역세권을 중심으로 도시를 개발하여 유동인구 수를 늘리는 TOD 개발 방식이 화제이다. 이는 대중교통 거점(지하철역)을 중심으로 인근 지역을 개발하여 역세권의 유동 인구 수를 늘이는 개발 방식으로 도시를 활성화함과 동시에 환경을 보존할 수 있도록 한다. 또한 최근에는 도시재생을 목적으로 역세권 개발 사업의 역할이 화제가 되고 있다. 하지만 과거 역세권 개발 사례를 살펴보면 특정 지역만 사업을 진행해 지역 간의 형평성 문제가 제기되어 왔다.

이에 본 프로젝트에서는 서울시 역세권 지역의 토지 유형별 TOD 균형 개발 방식을 제안하고자 한다. 서울시 지하철 네트워크 분석을 통해 중심성이 가장 낮은 5개의 역을 선별한다. 클러스터링을 수행하여 토지 유형별로 역세권을 분류한다. 군집 별로 주성분분석과 회귀분석을 진행하여 특정 토지 유형에서 가장 중요한 토지특성 및 시설 변수를 확인한다. 주성분을 독립변수로 두고 역세권의 유동인구를 예측하는 회귀모델을 구축하여 역세권의 대중교통 유동 인구를 늘리기 위한 방안을 제시하고자 한다.



연구 배경



TOD(Transit Orientation Development) 대중교통 지향형 개발

- 중심부(고밀지역)에는 대중교통 거점(지하철역)을 중심으로 대중교통 환승 센터와 상업 시설 등의 개발을 통해 지하철 역 주변의 유동인구가 증가하도록 대중교통 근접권의 도시를 활성화하고, 외곽지역은 개발을 지양하고 자연 생태 환경을 보존하는 형태
- 최근 지하철역 역세권 활성화사업으로 대중교통이 편리한 역세권의 고밀 복합개발을 통해 도시 내 주택공급, 균형발전을 촉진하겠다는 계획이 논의되고 있다

탄력받는 ‘역세권 개발’... 형평성 문제는 우려



- 도시를 재생하고자 하는 목적의 역세권개발사업이 화제이다
- 그러나, 과거 역세권 개발이 특정 지역에만 치우쳐 형평성 문제와 개발활성과 지역 vs 비활성화 지역의 개발격차를 형성하는 등의 문제를 일으켰다
- 서울시의 경우 강남 강북의 균형 발전, 역세권 공공시설 공급 불균형, 지하철역 역세권 지역의 도시 균형개발 등이 화제이다

RQ: 서울시 역세권을 균등하게 활성화시키기 위해 대중교통(지하철) 이용자수를 늘리는 TOD 방식을 제안한다

주제	저자	고찰/시사점/한계점
I. 지하철 2호선 역세권 별 유출입 특성과 역세권 토지이용에 관한 연구	김용민, 김한배, 정혜영, 이명훈 (2017).	지하철 2호선의 역세권 별 지하철 이용객의 이용패턴을 유형화하고 각 유형별 토지 이용 특성을 분석하여 역세권을 개발할 수 있는 방안을 도출하였다. ⇒ 지하철 이용 특성을 유형화할 때 클러스터링 등의 분석적 기법을 사용하지 않고 주관적인 해석에 따라 연구를 진행한 점이 아쉽다.
II. 클러스터링을 통한 역세권 잠재 개발역량 평가 연구: 서울시 2호선 대중교통 용량 및 토지 특성을 중심으로	이우섭, 강민희, 송재인, 황기연. (2021).	지하철 2호선 역세권의 대중교통 최대용량과 개발공시지가 등을 토대로 개발용량을 산출하였다. 클러스터링을 통해 역세권을 유형화하고 각 군집의 잠재적 개발 역량을 평가하였다. ⇒ 연구 대상을 2호선으로 한정된 점이 아쉽다.
III. 도시철도 이용률과 역세권 토지이용특성 간 연관성 분석 (광주광역시 사례로)	김형민, 정봉현. (2017).	광주광역시 지하철 1호선을 사례로 역세권 별 토지의 지역 용도 및 특성과 도시철도 이용률 간의 연관성을 파악하였다. ⇒ 통행 목적지의 역세권 범위를 정밀하게 반영되지 못한 점이 아쉽다. 또한 역세권 내 토지용도만을 고려하고 개발밀도 등 기타 여건을 반영하지 못한 것이 한계점이다.
IV. TOD(대중교통중심형) 도시개발 특성 및 발전전략에 관한 연구	박천보. (2016).	TOD형 도시개발의 유형을 제시하였고 국외의 TOD형 도시개발 성공 사례의 특성과 발전 전략을 고찰함으로써 국내의 TOD형 도시개발 방향을 제시하였다. ⇒ 국외 TOD형 성공 사례들을 참고하여 정형적인 연구를 이어가지 않고 그들의 TOD형 도시개발 전략을 모방하는 것에 그친 점이 아쉽다.
V. 사회 네트워크 분석을 이용한 지하철역 네트워크 구조 분석	한경훈, 서영수, 박근병. (2015).	지하철역을 노드, 승객들의 역간 이동을 엣지로 간주하여 네트워크 분석을 하였으며 중심성 지표를 활용하여 지하철역의 네트워크적 특성에 대해 고찰하였다. ⇒ 서울지하철 1~4호선만을 대상으로 네트워크를 구축했기 때문에 서울 전체 지하철역 네트워크를 나타낸다고 보기 어렵다. 또한 네트워크 특성을 단순히 분석한 것에서 연구가 그쳤다는 점이 아쉽다.

➡ 선행연구를 리뷰한 결과 연구 범위가 넓지 않다는 것을 확인했다. 또한 유동인구 수를 예측하는 모델을 구축하는 과정에서 토지특성 정보 외의 역세권 인근의 시설 데이터를 이용한 사례가 없다는 것을 확인했다.

본 프로젝트의 차별성은 다음과 같다.

- 1) 서울시 지하철역 전체로 연구 대상의 범위를 확대하여 연구를 진행하였다.
- 2) 토지특성정보와 역세권 인근 시설 데이터를 사용함으로써 독립변수의 다양성을 증가시켜 연구의 정확성을 높였다.



데이터 출처

SEOUL TOPIS 에서 제공하는 '대중교통 O/D 현황'
(Origin: 출발지, Destination: 도착지) 중 '역별 O/D' 활용
: 대중교통 O/D 현황 | 서울시 교통정보 시스템 - TOPIS (seoul.go.kr)

시간적 범위

2022년 04월 30일

공간적 범위

9개 노선 : 지하철 1 - 9 호선

사용한 변수

변수명	변수 유형	변수 설명
승차_역	Categorical	-
하차_역	Categorical	-
총_승객수	Numerical	일반 + 어린이 + 청소년 + 대학생 + 경로 + 장애인 + 국가유공자 + 직원 + 외국인경로

데이터 출처

국가교통 데이터 오픈마켓 에서 제공하는 '역세권 토지건물 특성정보' 활용
: 국가교통 데이터 오픈마켓 (bigdata-transportation.kr)

시간적 범위

2021년 12월

공간적 범위

역세권: 지하철 역사로부터 500m 이내의 부동산

사용한 변수

변수명	변수 유형	변수 설명
NEAR_SUBW_NM	Categorical	인접한 지하철역명
TOTAR	Numerical	연면적(m ²)(건축물대장 기준)
SIAR	Numerical	대지면적(m ²)(건축물대장 기준)



Introduction

Literature Review

Data Description

Methodology

Conclusion

References

데이터 출처

국가교통 데이터 오픈마켓 에서 제공하는 '공동주택
실거래정보' 활용

: 국가교통 데이터 오픈마켓 (bigdata-transportation.kr)

시간적 범위

2021년 04월

공간적 범위

역세권: 지하철 역사로부터 500m 이내의 부동산

사용한 변수

변수명	변수 유형	변수 설명
NEAR_SUBBW_NM	Categorical	인접한 지하철역명
BLDG_USE_NM	Categorical	건물 주용도명(건축물대장 기준) : 단독주택, 미등록, 공동주택, 제1종근린생활시설, 제2종근린 생활시설, 업무시설, 의료시설, 노유자시설, 종교시설, 교육연 구시설, 자동차관련시설, 창고 시설, 문화및집회시설, 공장, 위 험물저장및처리시설 등으로 구 분됨

데이터 출처

국가교통 데이터 오픈마켓 에서 제공하는 '역세권
토지용도 분포정보' 활용

: 국가교통 데이터 오픈마켓 (bigdata-transportation.kr)

시간적 범위

2017년 01월부터 2021년 01월까지

공간적 범위

역세권: 지하철 역사로부터 500m 이내의 부동산

사용한 변수

변수명	변수 유형	변수 설명
SUBSTAREA_NM	Categorical	지하철 역세권명
REFYM	VARCHAR	지표 기준년월
OALP_AVG	Numerical	시점수정 공시지가(원/m ²)의 평균
RESDT_AREA_RT	Numerical	전체 대비 주거용 토지면적 비율
COMRC_AREA_RT	Numerical	전체 대비 상업업무용 토지면적 비율
MIXBLDG_AREA_RT	Numerical	전체 대비 주상복합용 토지면적 비율
INDST_AREA_RT	Numerical	전체 대비 공업용 토지면적 비율
ETC_AREA_RT	Numerical	전체 대비 기타(전, 담, 임야) 토지면 적 비율

데이터 출처

서울 열린데이터 광장 에서 제공하는
'서울시 지하철 호선별 역별 시간대별
승하차 인원 정보' 활용

: 서울시 지하철 호선별 역별 시간대별 승하차 인원
정보> 데이터셋> 공공데이터 | 서울열린데이터광장
(seoul.go.kr)

시간적 범위

2021년 12월

공간적 범위

지하철 1 - 9 호선과 서울시 관할 운송기관
중 서울시에 해당하는 역

사용한 변수

변수명	변수 유형	변수 설명
지하철역	Categorical	-
승차인원	Numerical	1시간 간격으로 시간대 별 승차인원이 기록되 어 있음
하차인원	Numerical	1시간 간격으로 시간대 별 하차인원이 기록되 어 있음

Methodology

1 Preprocessing

프로젝트에 필요한 데이터셋을 새롭게 구축하고 목적에 맞게 **전처리**를 수행한다

2 SNA

서울시 지하철 네트워크에서 각 역의 **연결중심성**과 **아이겐벡터 중심성**을 확인하여 중심성이 낮은 역들을 확인하여 개발이 우선적으로 필요한 역들을 파악한다

3 Clustering

역세권을 **토지 유형 별로 분류**하기 위해 토지 용도별 면적 비율 변수들을 사용하여 **K-means 클러스터링**한다

4 PCA

클러스터 별 주성분 분석을 수행함으로써 토지 유형에 따라 어떠한 **변수의 중요도**가 높은 지 확인하고, **차원을 축소**한다

5 Regression

클러스터 별로 Y: 지하철 유동인구 수, X: 주성분(step4에서 구한 것)으로 하여 **다중회귀분석**을 수행한다. 주성분들의 **회귀 계수를 비교**하여 지하철 수요량에 많은 영향을 미치는 변수들을 확인하고자 한다.



1 Preprocessing

- 4개의 raw data set을 지하철역을 기준으로 **join** 시켜서 새로운 데이터프레임을 만들었다.
- 유의미한 값을 갖지 않는 칼럼은 제거하고 모든 변수에 대해 **정규화**했다.
- **상관성 분석**을 통해 시간대별 승차 하차 인원은 높은 연관성이 있다는 것을 확인하였고 추후 연구의 편리성을 위해 시간대별 승차인원을 다 합쳐서 승차인원이라는 **새로운 변수**를 만들어 **차원을 축소**하였다.

변수명	변수 유형	변수 설명
SUBSTATAREA_NM	Categorical	지하철 역세권명
승차인원	Numerical	1시간 간격으로 시간대별 승차인원이 기록되어 있음
하차인원	Numerical	1시간 간격으로 시간대별 하차인원이 기록되어 있음
FAR	Numerical	용적률(연면적/대지면적, TOTAR/SIAR)
REFYM	Varchar	지표 기준년월
OALP_AVG	Numerical	시점수정 공시지가(원/m ²)의 평균
RESDT_AREA_RT	Numerical	전체 대비 주거용 토지면적 비율
COMRC_AREA_RT	Numerical	전체 대비 상업업무용 토지면적 비율
MIXBLDG_AREA_RT	Numerical	전체 대비 주상복합용 토지면적 비율
INDST_AREA_RT	Numerical	전체 대비 공업용 토지면적 비율
ETC_AREA_RT	Numerical	전체 대비 기타(전, 답, 임야) 토지면적 비율
시설	Categorical	건물 주용도명(건축물대장 기준), 36개의 변수가 있음



2 SNA

- 서울시 지하철역 네트워크는
Node: 지하철역명, Edge: 승객의 이동 경로, Weight: Node i에서 j로 이동한 승객 수로 구성된 Undirected Network이다.
- 서울에서 상대적으로 역세권이 활성화되지 않은 지역을 선별하기 위해 네트워크에서 중심성이 낮은 역을 확인한다.

상관성 분석

	indeg	outdeg	eigen
indeg	1.000000	0.997450	0.988616
outdeg	0.997450	1.000000	0.986348
eigen	0.988616	0.986348	1.000000

연결 중심성은 in방향과 out방향에 따라 in deg와 out deg로, 아이겐벡터 중심성은 eigen으로 칼럼을 명하였다. 모든 역의 in deg, out deg, eigen 값을 확인한 결과 세 지표가 매우 높은 상관성을 보인다.

하위 노드 선별

서강대	0.222222	0.234921	0.017472
신답	0.228571	0.230159	0.017759
신내	0.290476	0.307937	0.023450
삼양사거리	0.295238	0.307937	0.025173
삼양	0.300000	0.279365	0.025121

<in deg 기준 하위 5개 노드>

신답	0.228571	0.230159	0.017759
서강대	0.222222	0.234921	0.017472
삼양	0.300000	0.279365	0.025121
둔촌오륜	0.303175	0.282540	0.023588
솔밭공원	0.314286	0.304762	0.026524

<eigen 기준 하위 5개 노드>

서강대	0.222222	0.234921	0.017472
신답	0.228571	0.230159	0.017759
신내	0.290476	0.307937	0.023450
둔촌오륜	0.303175	0.282540	0.023588
도림천	0.317460	0.306349	0.024144

<out deg 기준 하위 5개 노드>

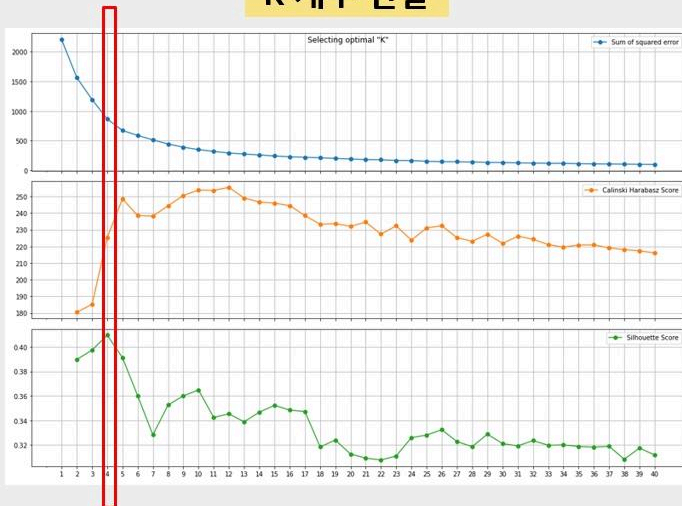
공통적으로 중심성 값이 낮은 역은 서강대, 신답, 신내, 삼양, 둔촌오륜인 것으로 파악된다.



3 Clustering

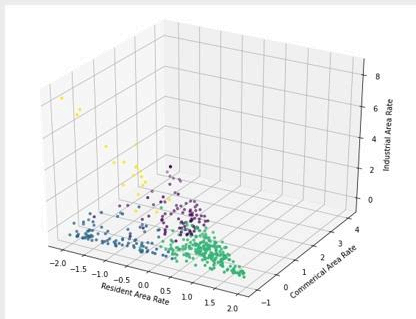
- RESDT_AREA_RT(주거면적비율), COMRC_AREA_RT(상업업무면적비율), MIXBLDG_AREA_RT(주상복합면적비율), INDST_AREA_RT(공업면적비율), ETC_AREA_RT(기타면적비율)를 변수로 하여 클러스터링한다.

K 개수 선별



Calinski Harabasz Score가 급격히 증가하고
Silhouette Score가 가장 높을 때를 선택하여 군집 수는
4개로 정한다.

시각화



전체 평균 실루엣 계수: 0.497
군집별 평균 실루엣 계수: 0.318237, 0.557685, 0.548900, 0.434902
⇒ 군집이 유의미하게 분류되었다고 판단된다.

결과 해석

	RESDT_AREA_RT	COMRC_AREA_RT	MIXBLDG_AREA_RT	INDST_AREA_RT	ETC_AREA_RT
k_means_cluster					
0	-0.682075	1.496502	0.525967	-0.200432	-0.506241
1	-1.068093	-0.783375	-0.745000	-0.073361	1.727509
2	0.761816	-0.247976	0.103551	-0.216058	-0.436827
3	-1.047940	-0.226819	-0.169637	4.025456	-0.559238

Cluster0: 주거용 토지면적 비율은 낮고, 상업용 토지면적 비율과 주상복합용 토지면적은 높은 지역이므로 **상업지구**이다. (91개)

Cluster1: 주거용, 상업용, 주상복합용 토지면적 비율은 낮고, 기타용 토지면적 비율이 높은 지역이므로 **녹지**이다. (93개)

Cluster2: 주거용 토지면적 비율이 높고, 기타용 토지면적 비율이 낮은 지역으로 **주거지구**이다. (238개)

Cluster3: 주거용 토지면적 비율이 낮고, 공업용 토지면적 비율이 높은 지역이므로 **공업지구**이다. (19개)



4 PCA

- 클러스터 별로 각 클러스터에서 **중요한 변수들을 선정**하고 **차원을 축소**하기 위해 주성분 분석을 수행한다.
- 주성분 분석은 데이터에서 패턴을 찾아내고 데이터의 유사성과 차이점을 강조하는 기법으로 클러스터 별로 지하철 유동인구에 가장 많은 영향을 끼치는 **중요 변수**를 알아내고자 하는 프로젝트의 목적과 잘 부합한다.
- 적절한 주성분의 개수는 아래의 **3가지 기준**을 모두 종합적으로 고려하여 선택하였다.
 - 1) Eigen value가 1이상일 때까지
 - 2) Scree plot 그래프 기울기가 완만해지기 전의 값까지
 - 3) 약 70% 이상의 분산을 설명하는 주성분 개수 채택

5 Regression

- 클러스터 별(토지 특성 유형에 따라)로 Y(**지하철 유동인구**)와 X(**토지용도비율, 용적률, 시설정보**)와의 관계를 나타내는 다중회귀모델을 구축한다.
- 주성분을 다중회귀분석의 X 변수로 사용함으로써 **다중공선성** 문제를 **해결**하였다.



전처리

- X 변수 중 '시설'은 특정 시설의 개수를 **100개 이상** 갖는 역세권이 존재할 때만 변수로 포함하였다. (0 값을 갖는 observation이 다수이므로 유의미한 차이를 나타나는 경우에만 선택하도록 한 것이다.)
- 모든 변수에 대해 **정규화**를 진행하여 주성분의 분산량의 왜곡이 없도록 하였다.

Cluster0(상업지역)

[PCA 데이터]

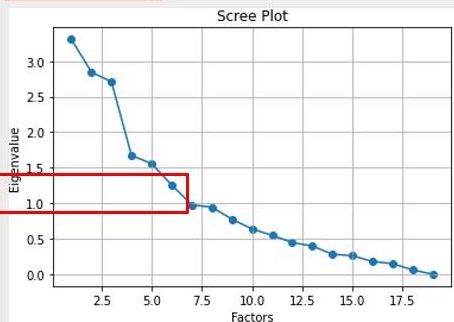
4 PCA

- X 변수: FAR, 단독주택, 공동주택, 제1종 근린생활시설, 제2종 근린생활시설, 업무시설, 교육연구시설, 자동차관련시설, 문화및집회시설, 공장, 판매시설, 숙박시설, 운수시설, 판매및영업시설, RESDT_AREA_RT, COMRC_AREA_RT, MIXBLDG_AREA_RT, INDST_AREA_RT, ETC_AREA_RT

주성분 개수 선정

eigen value: [3.30862272e+00 2.84781634e+00 2.70882435e+00 1.67269291e+00 1.55634776e+00 1.25582820e+00 9.75126743e-01 9.44273189e-01 7.74344274e-01 6.36760275e-01 5.47239038e-01 4.47004904e-01 3.96165967e-01 2.82027994e-01 2.59119188e-01 1.79433201e-01 1.48461121e-01 5.99097102e-02 2.12065289e-06]

적합한 PC 개수: 6



누적 분산 설명 비율:

[0.17413804 0.32402311 0.46659281 0.55462928 0.63654232 0.70263854 0.753951 0.80365959 0.84441455 0.87792825 0.90673031 0.93025688 0.95110772 0.9659513 0.97958915 0.989033 0.99684675 0.99999989 1.]

주성분 설명

	PC1	PC2	PC3	PC4	PC5	PC6
FAR	0.259131	0.350477	0.102116	-0.042335	-0.031629	0.357948
단독주택	0.260522	-0.083948	-0.390445	-0.137629	-0.122868	-0.205859
공동주택	0.171625	-0.064302	-0.443184	-0.132975	-0.139738	0.198999
제1종근린생활시설	0.433216	-0.124681	0.029161	-0.112974	0.033729	-0.296950
제2종근린생활시설	0.450144	-0.139473	0.004402	0.193625	0.217692	-0.039320
업무시설	0.311148	-0.036764	0.129774	-0.133492	-0.130886	0.345529
교육연구시설	0.274249	-0.039501	-0.035850	0.381560	0.238494	-0.064673
자동차관련시설	0.022839	-0.014897	-0.042866	-0.111447	0.139992	0.626556
문화및집회시설	0.221088	-0.098715	0.036475	0.409683	0.419247	0.128308
공장	0.009315	0.007453	-0.029833	-0.447241	0.500053	-0.150562
판매시설	0.286085	0.172996	0.136188	-0.066931	-0.121599	-0.191418
숙박시설	0.235960	-0.086951	0.024755	-0.298363	-0.208266	-0.052158
운수시설	0.109605	0.537823	0.005640	0.027617	0.045785	0.031359
판매및영업시설	0.059795	0.540207	0.011137	0.014682	0.007621	0.004545
RESDT_AREA_RT	-0.047629	-0.142173	-0.495081	0.093261	-0.000162	0.197682
COMRC_AREA_RT	0.105306	-0.169883	0.521936	-0.114550	-0.125713	0.035343
MIXBLDG_AREA_RT	0.097356	0.342661	-0.278710	0.044847	-0.132675	-0.206973
INDST_AREA_RT	-0.070794	0.083865	-0.055082	-0.439797	0.528477	-0.018091
ETC_AREA_RT	-0.206705	0.161132	0.010370	0.236446	0.153007	-0.145032

PC1: 근린생활시설,

업무시설, 판매시설

PC2: 운수 및 영업 시설

PC3: 주거지 및 상업 발달 정도

PC4: 교육 및 문화 시설

PC5: 공업 시설

PC6: 자동차 관련 시설

적절한 주성분의 개수는 6개



Cluster0(상업지역)

4 PCA

주성분 점수

	PC1	PC2	PC3	PC4	PC5	PC6
SUBSTATAREA_NM						
가락시장역	-1.293041	-0.262838	0.667073	0.126122	-0.440613	0.397877
강남역	1.232640	-0.045258	1.433354	0.754810	0.006096	1.624707
강동역	0.222596	0.163411	-3.359534	-0.290943	-1.232015	0.420299
경복궁(정부서울청사)역	-0.268306	-0.603687	-0.907915	0.589654	-0.082073	-0.374604
고속터미널역	-0.507038	0.473211	0.093104	2.010392	0.885148	0.882561
...
한대앞역	-2.458247	0.496375	-0.336428	1.421206	0.440381	-1.091230
한성백제역	-1.571091	-0.481875	-0.365818	0.338661	-0.427159	0.286555
합정역	0.614244	-0.340264	-1.851990	0.231957	-0.508437	0.127117
홍대입구역	5.065308	3.194251	-0.497549	1.030718	-0.032504	0.686175
회현(남대문시장)역	2.544404	0.557084	2.280536	0.087998	-0.552005	-1.460164

91 rows × 6 columns

주성분 점수를 통하여 각 역세권에 발달한 시설 정보와 주거지 및 상업 발달 정도를 파악할 수 있다.

Cluster0(상업지역)

5 Regression

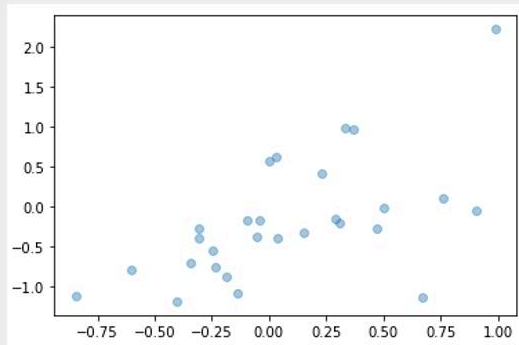
[Regression 데이터]

- Y 변수: 지하철 유동인구(승차인원+하차인원)
- X 변수: PC1, PC2, PC3, PC4, PC5, PC6

다중회귀모델

$$Y = 0.08038602 + 0.16055849 * PC1 - 0.06265025 * PC2 + 0.02995336 * PC3 - 0.00857878 * PC4 - 0.04755446 * PC5 + 0.18491758 * PC6$$

모델 성능평가



R2 = 0.225

RMSE = 0.654



Cluster1(녹지)

4 PCA

[PCA 데이터]

- X 변수: FAR, 단독주택, 공동주택, 제1종 근린생활시설, 제2종 근린생활시설, 업무시설, 교육연구시설, 공장, 동식물관련시설, 판매시설, 교정및군사시설, 관광휴게시설, RESDT_AREA_RT, COMRC_AREA_RT, MIXBLDG_AREA_RT, INDST_AREA_RT, ETC_AREA_RT

주성분 설명

PC1: 근린생활시설과 개발밀도

PC2: 연구 및 군사 시설

PC3: 공업 발달 정도

PC4: 상업 발달 정도

PC5: 판매 시설

Cluster1(녹지)

5 Regression

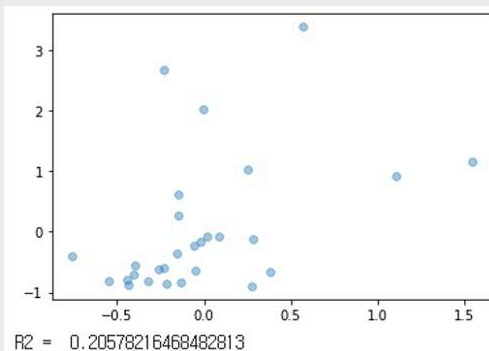
[Regression 데이터]

- Y 변수: 지하철 유동인구(승차인원+하차인원)
- X 변수: PC1, PC2, PC3, PC4, PC5

다중회귀모델

$$Y = -0.01537099 + 0.18375623 * PC1 - 0.09339608 * PC2 + 0.04996996 * PC3 + 0.11186993 * PC4 + 0.10600636 * PC5$$

모델 성능평가



R2 = 0.206

RMSE = 0.984



Cluster2(주거지역)

4 PCA

[PCA 데이터]

- X 변수: FAR, 단독주택, 공동주택, 제1종 근린생활시설, 제2종 근린생활시설, 업무시설, 종교시설, 교육연구시설, 공장, 판매시설, 숙박시설, 근린생활시설, RESDT_AREA_RT, COMRC_AREA_RT, MIXBLDG_AREA_RT, INDST_AREA_RT, ETC_AREA_RT

주성분 설명

- PC1: 근린생활시설과 업무시설
 PC2: 공업 발달 정도
 PC3: 주거지 및 녹지 발달 정도
 PC4: FAR(용적률)
 PC5: 상업 발달 정도

Cluster2(주거지역)

5 Regression

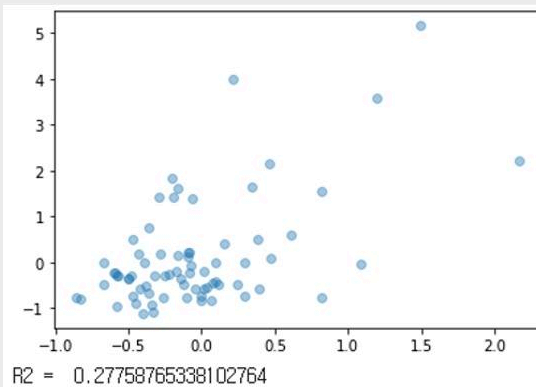
[Regression 데이터]

- Y 변수: 지하철 유동인구(승차인원+하차인원)
- X 변수: PC1, PC2, PC3, PC4, PC5

다중회귀모델

$$Y = -0.05200192 + 0.2304288 * PC1 + 0.03553716 * PC2 \\ - 0.01388135 * PC3 + 0.0211828 * PC4 + 0.15218891 * PC5$$

모델 성능평가



R2 = 0.278

RMSE = 0.994



Cluster3(공업지역)

4 PCA

[PCA 데이터]

- X 변수: FAR, 단독주택, 공동주택, 제1종 근린생활시설, 제2종 근린생활시설, 교육연구시설, 창고시설, 문화및집회시설, 공장, 숙박시설, 운수시설, RESDT_AREA_RT, COMRC_AREA_RT, MIXBLDG_AREA_RT, INDST_AREA_RT, ETC_AREA_RT

주성분 설명

PC1: 주거지와 근린생활시설 (the smaller, the better)

PC2: 연구 및 창고 시설

PC3: 상업 및 주거용 토지 비율

PC4: 여행 산업 발달 정도

Cluster3(공업지역)

5 Regression

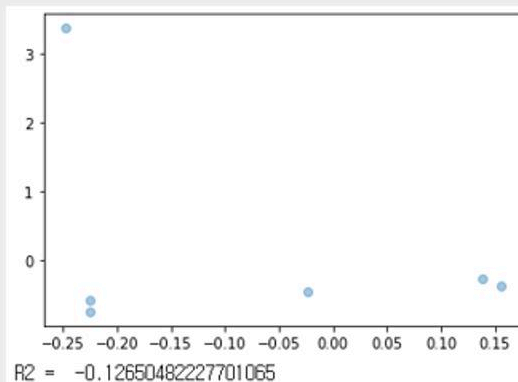
[Regression 데이터]

- Y 변수: 지하철 유동인구(승차인원+하차인원)
- X 변수: PC1, PC2, PC3, PC4, PC5

다중회귀모델

$$Y = -0.07115131 - 0.07743486 * PC1 - 0.08164706 * PC2 - 0.00513735 * PC3 + 0.13396244 * PC4$$

모델 성능평가



$$R^2 = -0.127$$

$$RMSE = 1.537$$

클러스터3에 해당하는 데이터 수가
19개로 학습시키는 데에 적절하지 않음.
R2 값도 -0.127로 모델을 신뢰할 수 없음.



TOD 균형 개발 방안 예시 -

서강대역

	PC1	PC2	PC3	PC4	PC5
서강대역	-0.105659	-0.092924	-0.995124	1.502398	0.966967

PC1: 근린생활시설과 업무시설

PC2: 공업 발달 정도

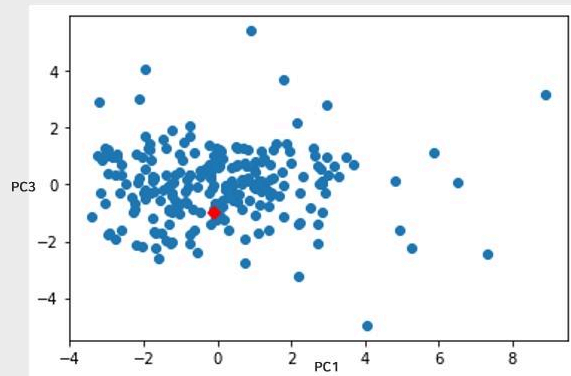
PC3: 주거지 및 녹지 발달 정도

PC4: FAR(용적률)

PC5: 상업 발달 정도

$$Y = -0.05200192 + 0.2304288 * PC1 + 0.03553716 * PC2 - 0.01388135 * PC3 + 0.0211828 * PC4 + 0.15218891 * PC5$$

- 서강대역은 클러스터2(주거지역)에 속하므로 **클러스터2의 회귀모델**을 이용하여 분석.
- 제1주성분의 회귀계수가 가장 크고, 서강대역의 제1주성분 점수는 비교적 낮은 값을 가지므로, **근린생활시설**과 **업무시설**을 **확대**하는 것이 우선적.
- 제3주성분의 경우, 녹지 면적이 늘어날수록 Y(지하철 이용량)가 증가하므로 **공원**과 같은 **녹지**를 개발 해야함.



시사점

- 환경친화적인 TOD 방식을 활용하여 서울시 역세권의 균형 개발을 도모한다는 것에 의의가 있다.
- 역세권 주변 시설을 고려하여 대중교통 수요량을 예측했기 때문에 토지 특성 정보 외의 대중교통 수요량에 영향을 미치는 새로운 변수를 찾았다는 것에 의의가 있다.

한계점

- TOD 방식을 제안하는 데 연구 대상의 범위를 지하철로 한정된 점이 아쉽다. 역세권의 대중교통 이용량을 정확히 계산하려면 버스, 자전거 등의 이용률까지 고려해야 할 것이다.
- 단일 토지 유형 별 제안을 하는 모델을 구축했기 때문에 복합 개발을 목적으로 하는 경우는 해당 모델을 적용하기에는 한계가 있다. 따라서 복합 유형 개발에 대해 제언할 수 있는 후행 연구가 필요하다.



References

[논문]

- 김용민, 김환배, 정혜영, 이명훈. (2017). 지하철 2호선 역세권 별 유출입특성과 역세권 토지이용에 관한 연구. 한국지역개발학회지, 29(3), 157-179.
- 김한승욱. (2022). 일본 역세권 복합개발형 도시재생사업의 특성과 정책적 시사점에 관한 연구. 한국산학기술학회 논문지, 23(4), 541-548.
- 김형민, 정봉현. (2017). 도시철도 이용률과 역세권 토지이용특성간 연관성 분석. 한국지역개발학회 세미나 논문집, 1-11.
- 박천보. (2016). TOD(대중교통중심)형 도시개발 특성 및 발전전략에 관한 연구. 한국산학기술학회 논문지, 17(2), 635-641.
- 이우섭, 강민희, 송재인, 황기연. (2021). 클러스터링을 통한 역세권 잠재 개발역량 평가 연구 : 서울시 2호선 대중교통 용량 및 토지 특성을 중심으로. 국토계획, 56(6), 113-127.
- 한경훈, 서영수, 박근병. (2015). 사회 네트워크 분석을 이용한 지하철역 네트워크 구조 분석. 한국철도학회 학술발표대회논문집, (), 34-40.

[데이터 출처]

- 국가교통데이터의 국가교통 데이터 오픈마켓 (bigdata-transportation.kr)
- SEOUL TOPIS의 대중교통 O/D 현황 | 서울시 교통정보 시스템 - TOPIS (seoul.go.kr)
- 서울 열린데이터 광장의 서울시 지하철 호선별 역별 시간대별 승하차 인원 정보 > 데이터셋 > 공공데이터 | 서울열린데이터광장 (seoul.go.kr)



Thank you

and Q&A / Feedback

