

POI Embedding

Sundong Kim, BeomYoung Kim

Dec 13th, 2019

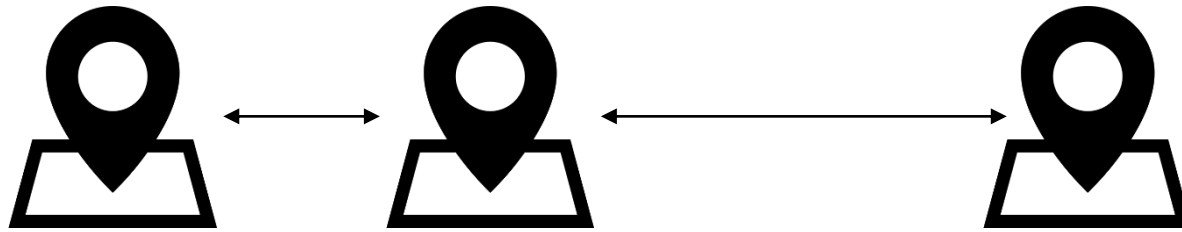
Embedding의 중요성

- 텍스트 기반의 모델 : 컴퓨터가 자연어를 이해하기 위해 숫자로 치환이 필요
 - One-hot vector
 - Embedding
- One-hot vector의 단점
 - 저장 공간 측면에서의 비효율
 - 단어 사이의 유사도 표현 불가능

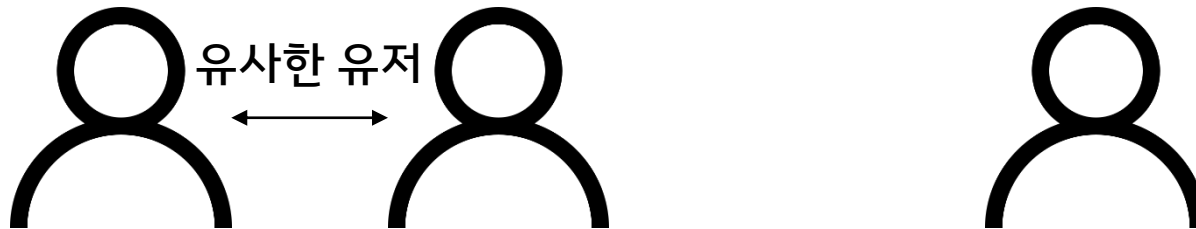
⇒ Poi data에서도 **카테고리, 브랜드 이름** 등 관계를 파악해야 할 텍스트가 존재

장소 Embedding의 활용

- 장소 Embedding을 통해 장소간의 유사도 수량화



- 특정 유저가 방문한 장소들의 Embedding을 통해 유저 Embedding을 정의하여 유저간의 유사도 측정
⇒ 타겟 마케팅 가능



Kaist_sample_final

제공받은 데이터의 주된 속성 (총 428044개의 logs)

- Adid : 특정 유저 id
- searchplace_response_place_name : 위치인식 된 매장 이름 (브랜드인 경우 브랜드 이름)
- searchplace_response_place_branch_name : 위치인식 된 매장의 지점 명
- searchplace_response_place_category : 위치인식 판정된 매장의 카테고리
- searchplace_response_place_addr_road : 위치인식 판정 결과의 도로명 주소
- searchplace_response_complex_name : 복합몰 이름
- searchplace_response_complex_branch_name : 복합몰 지점 이름
- searchplace_response_complex_category : 복합몰의 카테고리
- client_lat : 위치인식 로그를 남긴 기기의 위도
- client_lng : 위치인식 로그를 남긴 기기의 경도

결과물

- embedding+category.csv (level 1,2,3,4)
 - CATEGORY_LIST_232에 존재하는 level별 카테고리의 embedding
- embedding+searchplace_response_place_category.csv
 - Logs에 존재하는 카테고리의 embedding
- embedding+searchplace_response_place_name.csv
 - Logs에 존재하는 브랜드의 embedding
- embedding+poi.csv
 - Logs에 존재하는 (브랜드+지점)의 embedding
- embedding+complex.csv
 - Logs에 존재하는 (복합물+지점)의 embedding

Poincare Embedding

Poincare Embedding

- Loplatt data의 경우 장소가 카테고리, 브랜드, 특정 브랜드 지점 등 계층적 구조를 가지고 있음

⇒ 계층적 관계를 가지는 구조를 이해하는 Embedding 방법론이 필요

Poincare Embedding

- Poincare ball model
 - Hyperbolic geometry(쌍곡선 기하학)의 개념을 구현
 - 노든 간의 트리 구조(계층적 구조)를 잘 나타냄

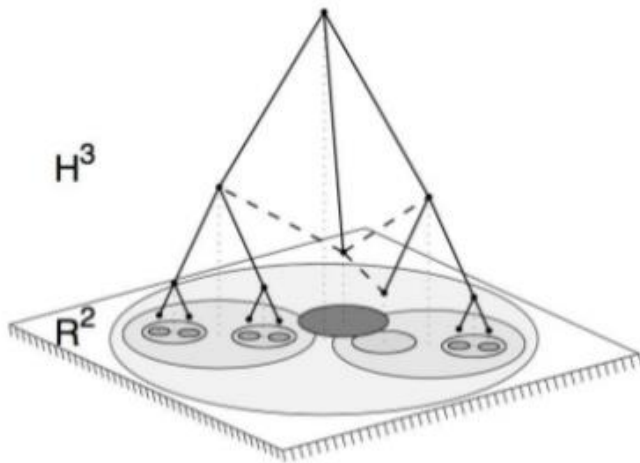


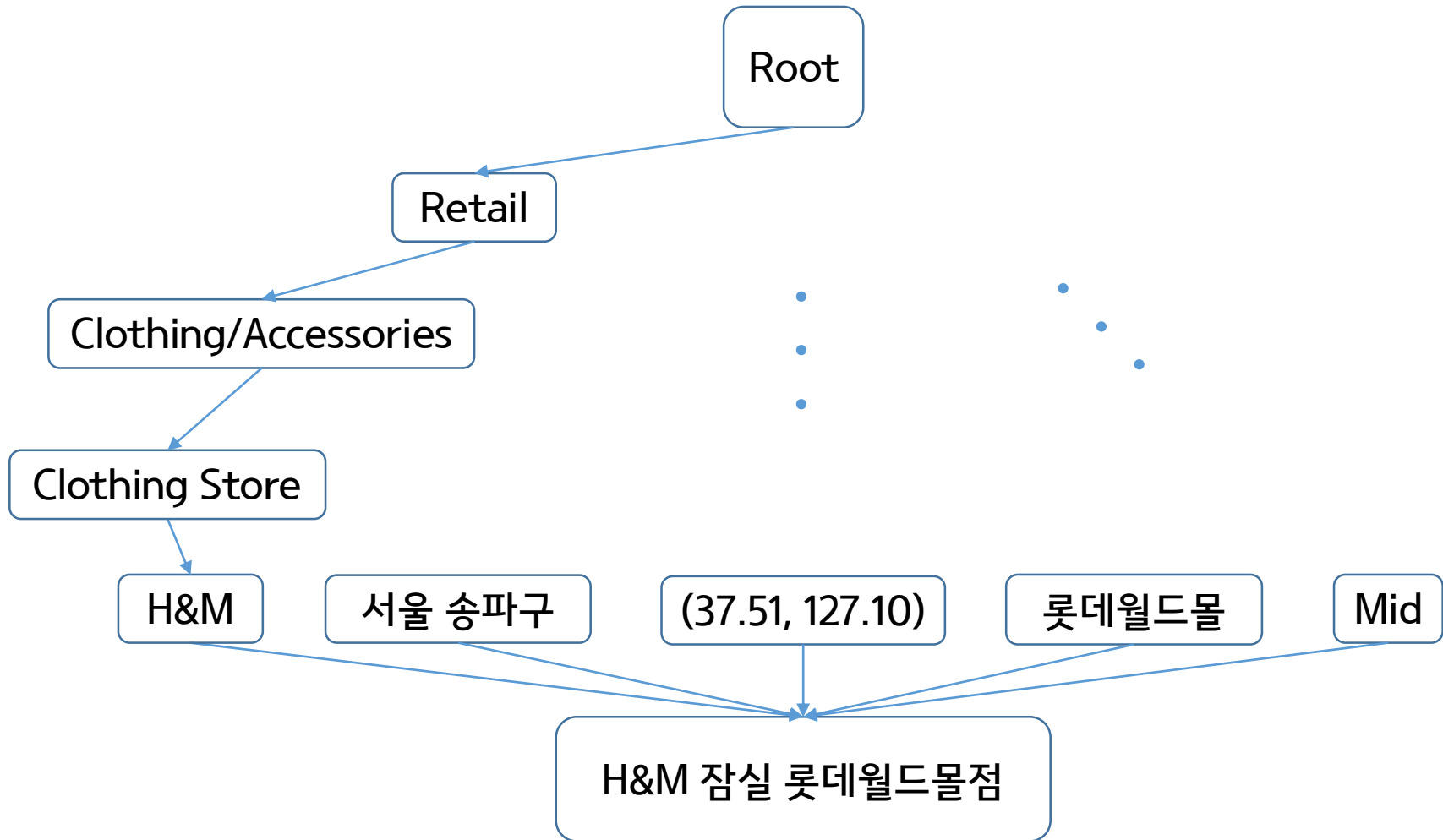
FIG. 2: Mapping between disks in the Euclidean plane \mathbb{R}^2 and points in the Poincaré half-space model of the three-dimensional hyperbolic space \mathbb{H}^3 [21]. The x, y -coordinates of disks in \mathbb{R}^2 are the x, y -coordinates of the corresponding points in \mathbb{H}^3 . The z -coordinates of these points in \mathbb{H}^3 are the radii of the corresponding disks. This mapping represents the tree-like hierarchy among the disks. Two points in \mathbb{H}^3 are connected by a solid link if one of the corresponding disks is the minimum-size disk that fully contains the other disk. This hierarchy is not perfect; thus, the tree structure is approximate. The darkest disk in the middle partially overlaps with three other disks at different levels of the hierarchy. Two points in \mathbb{H}^3 are connected by a dashed link if the corresponding disks partially overlap. These links add cycles to the tree. The shown structure is thus not strictly a tree, but it is hyperbolic [27].

FIG. 2 : <https://arxiv.org/pdf/1006.5169.pdf>

계층적 구조의 주요 속성

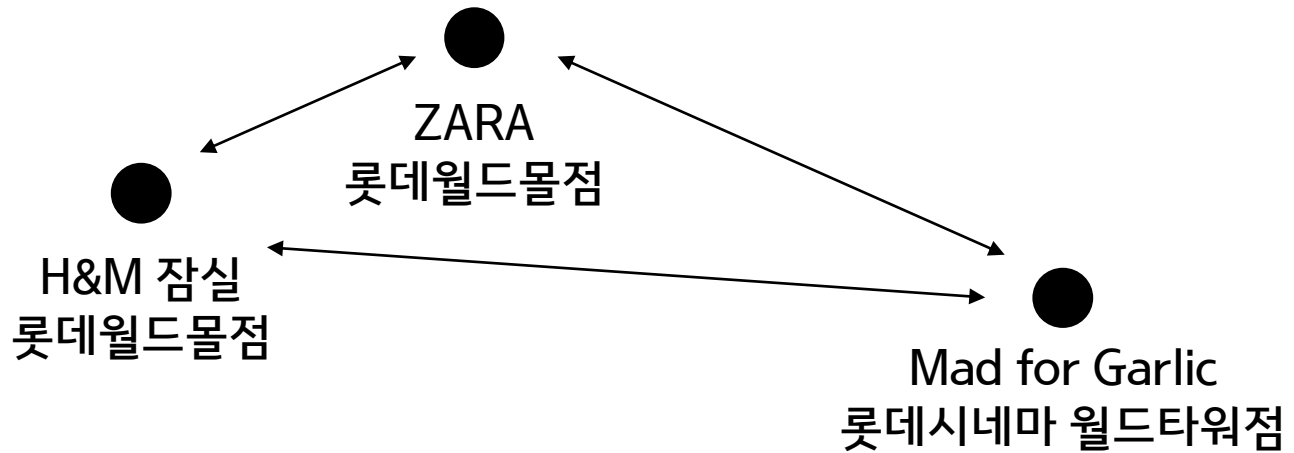
- 계층적 카테고리 와 브랜드
- 지리적 유사도
 - 도로명 주소 군집
 - 위, 경도를 이용한 군집
- 시간대별 효과
- 복합물의 효과

POI 계층 구조의 예시



Results & Analysis

가정



가까이 위치할수록, 비슷한 카테고리나 브랜드를 가질수록, 비슷한 시간대에 방문을 많이 했을 수록, 그리고 같은 복합몰에 위치할 수록 POI간의 유사도가 클 것이다.

예시

- 같은 복합물에 위치할 때, 비슷한 카테고리를 가지는 poi를 유사하다고 판단한다

| | | |
|-------------------|-------------------------------|----------------|
| Cosine similarity | Mad for Garlic 롯데시네마 월드타워점 | H&M Men 롯데월드몰점 |
| H&M 잠실 롯데월드몰점 | 0.0162 | 0.2811 |

| | | |
|-------------------------------|-------------|-------------|
| Cosine similarity | 서래냉면 롯데월드몰점 | 무인양품 롯데월드몰점 |
| Mad for Garlic 롯데시네마 월드타워점 | 0.5107 | 0.1915 |

예시

- 같은 부모를 가지는 카테고리 집합이 더 유사하다

| Cosine similarity | Snack Bar | Costmetics/Perfumes |
|-------------------|-----------|---------------------|
| Restaurants | 0.0649 | -0.0165 |

| Cosine similarity | Middle School | Pharmacy/Medical Equipments |
|-------------------|---------------|-----------------------------|
| College Education | 0.886 | 0.315 |

예시

- 복합몰이 고정되었을 때, 비슷한 브랜드나 카테고리가 유사하다
- 무인양품 롯데월드몰점과 '롯데월드몰점'에 존재하는 poi 사이의 Cosine similarity

| 무인양품 롯데월드몰점 | Cosine similarity |
|-----------------------|-------------------|
| Bills 잠실점 | 0.5470 |
| H&M Men 롯데월드몰점 | 0.5364 |
| THE NORTH FACE 롯데월드몰점 | 0.5297 |
| ... | ... |
| 삼보당호떡 롯데월드몰점 | -0.0764 |
| KON 롯데월드몰점 | -0.0862 |
| 모던하우스 롯데월드몰점 | -0.0990 |
| LAMEN'S 롯데월드몰점 | -0.1303 |

예시

- 장소가 고정되었을 때, 비슷한 브랜드나 카테고리가 유사하다
- 파리바게뜨 마포점과 '서울 서대문구'에 있는 poi 사이의 Cosine similarity

| 파리바게뜨 마포점 | Cosine similarity |
|--------------|-------------------|
| 파리바게뜨 충정디오빌점 | 0.6645 |
| 서대문족발 본점 | 0.6049 |
| 파리바게뜨 서대문행복점 | 0.5880 |
| 빵파네 서대문점 | 0.5736 |
| 롯데리아 서대문역점 | 0.5636 |
| 누베이크 서대문점 | 0.5622 |
| 투썸플레이스 임광빌딩점 | 0.5419 |
| ... | ... |

예시

- 브랜드나 카테고리가 고정되었을 때 가까운 poi가 유사하다
- Lotteria와 McDonald's 대전부사DT점 사이의 Cosine similarity

| 맥도날드 대전부사DT점 | Cosine similarity |
|---------------|-------------------|
| 롯데리아 대전은행점 | 0.6792 |
| 롯데리아 인천원당점 | 0.2485 |
| 롯데리아 홈플러스김해점 | 0.2343 |
| 롯데리아 한남대점 | 0.1971 |
| ... | ... |
| 롯데리아 서산호수공원점 | -0.3188 |
| 롯데리아 롯데백잠실광장점 | -0.3301 |
| 롯데리아 캐슬플자점 | -0.3586 |

예시

- 같은 브랜드의 효과와 지리적 효과 중 어떤 것이 우세한지 판단
- (브랜드가 파리바게뜨인 poi+같은 위,경도 군집에 있는 poi)와 파리바게뜨 한남대점 사이의 Cosine similarity
- 지리적인 영향이 조금 더 컸고, 대신에 파리바게뜨 사이에서는 가까울수록 유사하였음

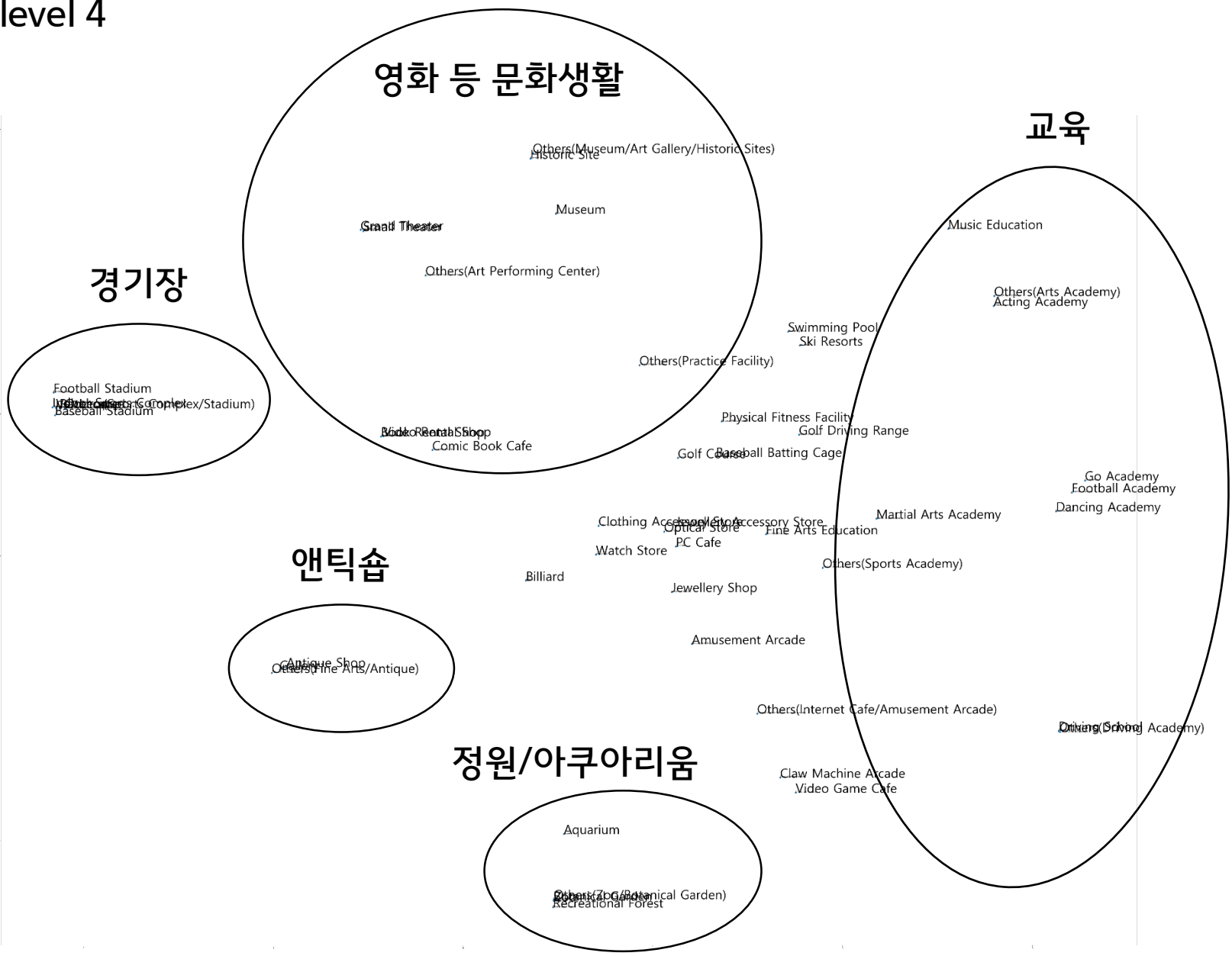
| 파리바게뜨 한남대점 | Cosine similarity |
|---------------|-------------------|
| 큰맘할매순대국 한남대점 | 0.8880 |
| 에빵버거 한남대점 | 0.8858 |
| 치킨과바람피자 한남대점 | 0.8837 |
| 차오차오 한남대점 | 0.8612 |
| 한마음 정육식당 한남대점 | 0.8418 |
| ... | ... |

| 파리바게뜨 한남대점 | Cosine similarity |
|--------------|-------------------|
| 파리바게뜨 대전역점 | 0.7646 |
| 파리바게뜨 대전효동점 | 0.5408 |
| 파리바게뜨 왕십리역사점 | 0.4891 |
| 파리바게뜨 화명점 | 0.4219 |
| ... | ... |

PCA & T-SNE

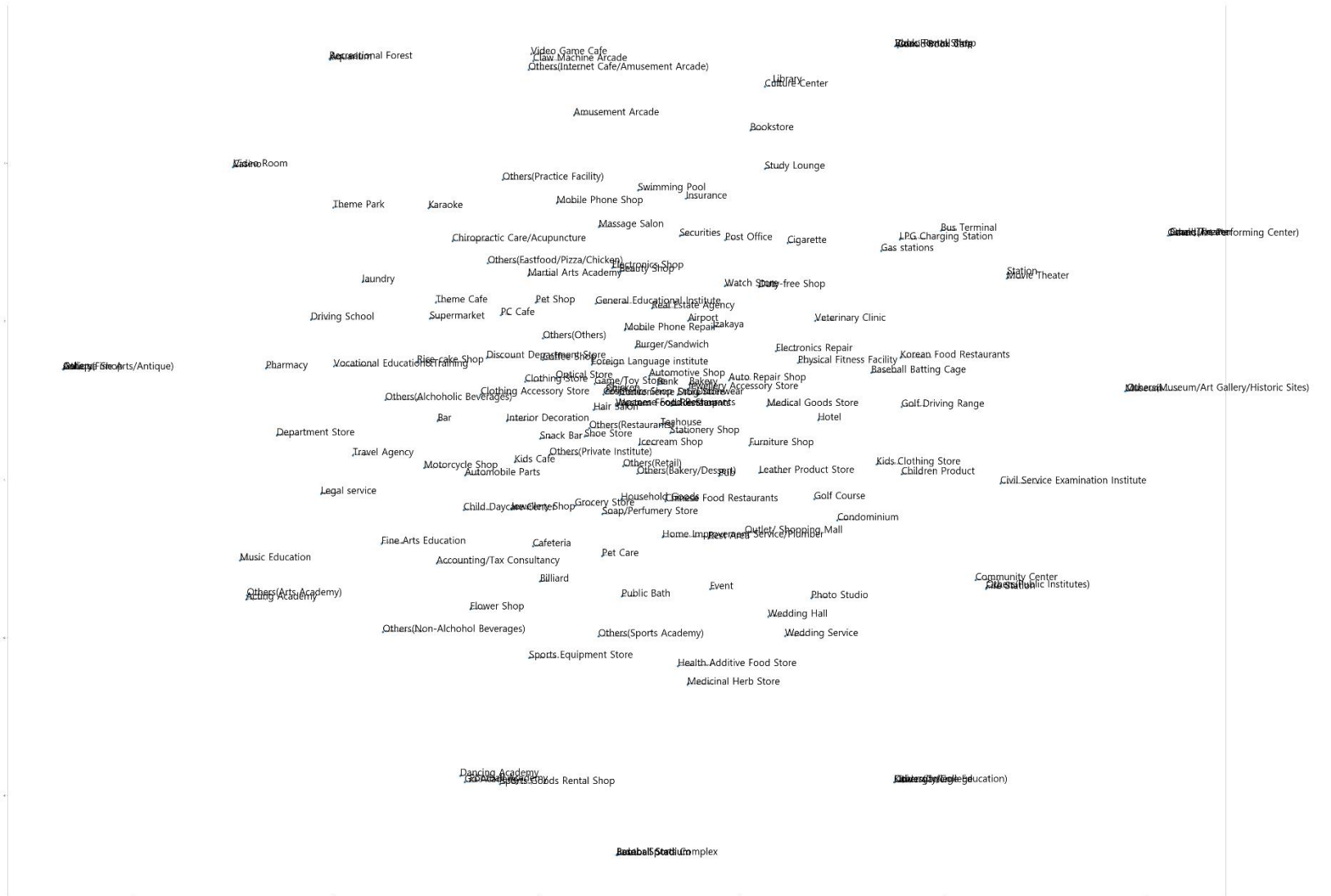
고차원의 벡터 차원 축소 및 시각화

Category level 4



searchplace_response

_place_category



Poi 중 일부(100개)

⇒ 가까운 지역이 모여 있음



Future work

Future Work

- 장소 Embedding을 더욱 세밀화
 - 결측값 처리가 필요함 : 장소가 미등록되어 위, 경도 정보만 가진 경우가 많음
 - 위, 경도 정보를 외부 데이터로 geo mapping하여 주소를 알아냄 (외부 데이터 이용)
 - 장소를 나타낼 더 자세한 특징이 필요함 : 위치나 시간대 정보를 더욱 세분화하여 복잡한 트리를 구성
- 유저 Embedding을 정의
 - 유저가 과거 방문했던 기록들을 Embedding하여 이용
 - 특정 장소에 방문한 횟수를 반영해야 할 것임