

# 차량 운전자의 내비게이션 사용 이력 데이터셋을 활용한 개인화 POI 추천 시스템

## A Personalized POI Recommender System Using Drivers' Navigation History

### 요 약

최근, 위치 기반 소셜 네트워크의 사용이 활발해지면서, 개인화 POI 추천에 관한 다양한 연구들이 산업계/학계에서 활발히 이뤄지고 있다. 하지만, 그 동안의 개인화 POI 추천 연구는 주로 위치 기반 소셜 네트워크 서비스 이용자들 (즉, 도보 또는 대중 교통을 통해 이동하는 사용자들)만을 대상으로 진행되어 왔다. 즉, 차량 운전자 대상의 개인화 POI 추천 연구는 진행이 미흡한 실정이었다. 따라서 본 연구에서는, 실제 차량 운전자들로부터 총 6개월동안 수집된 내비게이션 데이터셋을 이용해, 운전자가 방문 후 남긴 리뷰 및 그룹 별 인기도를 이용하여 개인화 POI 추천을 수행하는 방법을 새롭게 제안한다. 실세계 내비게이션 데이터셋을 이용한 실험을 통해, 본 연구에서 제안하는 방법이 인기도 기반의 단순 랭킹을 이용하는 비개인화 추천 방법 대비, 최대 약 15%까지 추천 정확도를 개선할 수 있음을 확인하였다.

### 1. 서 론

최근, 본인이 방문했던 장소들을 다른 사용자와 함께 공유하며 소통하는 위치 기반 소셜 네트워크 (LBSNs, Location Based Social Networks) 서비스들이 활발하게 사용되고 있다 (예: Foursquare, Gowalla). 대부분의 이러한 위치 기반 소셜 네트워크 서비스들은, 사용자가 자신이 방문했던 장소들에 남긴 체크인 기록을 바탕으로 해당 사용자가 흥미를 가질 만한 장소(POI, Point of Interest)를 추천하는 개인화 POI 추천 기능을 제공한다.

사용자들의 체크인 기록 및 장소들에 대한 만족도 정보 등을 바탕으로, LBSNs 도메인을 위한 POI 추천 기법들은 그동안 활발히 연구되어 왔다 [1-4]. 하지만, 이러한 연구들은 LBSNs 도메인의 특성 상, 주로 도보 또는 대중 교통을 통해 이동하는 사용자들만을 개인화 POI 추천의 고려 대상으로 삼는 경우가 많았다. 즉, 현재까지, 차량 운전자들 대상의 개인화 POI 추천을 수행하는 연구들은 진행이 매우 미흡한 상황이었다.

따라서, 본 연구에서는 실제 차량 운전자들로부터 총 6개월동안 수집된 내비게이션 이용 이력 데이터셋을 통해, 차량 운전자 대상의 개인화 POI 추천시스템을 새롭게 제안하고자 한다. 이를 위해 우리는 먼저, LBSNs 도메인을 타겟으로 제안된 기존 POI 추천 모델들의 주요 특징 및 수행 과정을 살펴본다 [1-4]. 다음, 이들 중 한 모델을 선택하여 [4], 운전자들의 내비게이션 이용 이력 데이터셋을

효과적으로 학습할 수 있도록, 내비게이션 도메인에 맞게 해당 모델을 적절히 수정한다. 우리는 이러한 모델을, 내비게이션 도메인을 위한 개인화 POI 추천 모델로서 새롭게 제안한다.

실세계 내비게이션 데이터셋을 이용한 실험을 통해, 제안하는 본 개인화 POI 추천 모델이 인기도 기반의 단순 랭킹을 이용하는 비개인화 POI 추천 방법 대비, 최대 약 15%까지 개선된 추천 정확도를 보일 수 있음을 확인하였다.

### 2. 관련 연구

LBSNs 도메인을 타겟으로 하는 기존의 POI 추천 시스템들은, POI의 위치 또는 사용자의 현 위치 등과 같은 지리적 정보를 공통적으로 사용한다 [1-4]. Ye [1]는 먼저, 사용자가 체크인 한 적 있던 POI들은 대부분 서로 가까운 거리 내에 위치하고 있음을 확인하였다. 이러한 관찰 결과를 기반으로, [1]은 USG 모델을 통해 거리에 따른 방문 확률을 예측함으로써, 개인화 POI 추천을 수행하였다. Zhao [2]에 의해 제안된 Geo-Teaser는 사용자의 체크인 시퀀스 데이터와 요일 별 체크인 경향, POI들 간의 거리 정보를 종합적으로 사용하여 요일 별 개인화 추천을 수행한다. Wang [3]은 POI의 위/경도 지리 정보 및 사용자의 체크인 데이터를 사용하여 사용자 선호도 및 지리적 특성을 고려한 추천 기법, GeoIE를 제안한다. 마지막으로, Ma [4]는 사용자의 체크인 데이터와 POI들 간의 거리 정보를 사용하여,

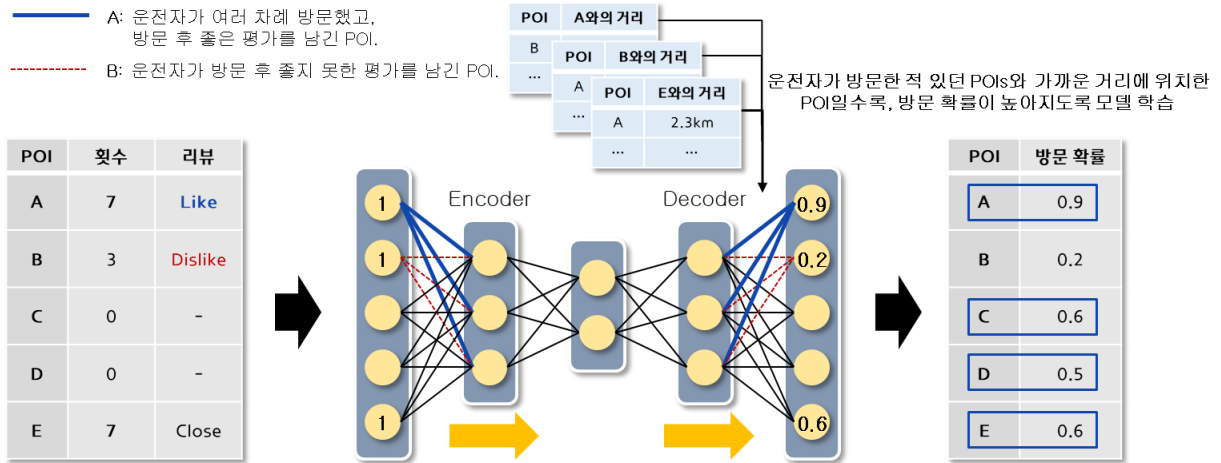


그림 1. 본 논문에서 제안하는 개인화 POI 추천 모델의 구조.

오토인코더 모델을 통해 사용자의 취향을 학습하는 개인화 POI 추천 방법, SAE-NAD를 제안한다.

## 2.1 SAE-NAD

오토인코더 모델 기반의 SAE-NAD는 크게 (1) Self-Attentive Encoder와 (2) Neighbor-Aware Decoder로 구분된다 [4]. (1)에서는, 사용자의 체크인 기록에서 관찰되는 각 POI의 중요도를 attention network를 통해 계산한 뒤, 중요도와 함께 각 POI를 인코딩 함으로써, 사용자 임베딩을 수행한다. 다음, (2)에서는, (1)에서 임베딩 된 사용자 벡터와 각 POI 간의 interaction을 구함으로써, 해당 사용자의 각 POI에 대한 방문 확률을 예측한다. 이 때, POI의 지리적 정보를 함께 고려하여, 사용자가 체크인했던 POI들과 가까운 거리에 위치한 POI 일수록, 방문 예측 확률이 더 높아지도록 계산된다.

## 3. 제안 방안

표 1. POI 카테고리의 종류 및 해당 항목

| 카테고리             | 항목                                                                          |
|------------------|-----------------------------------------------------------------------------|
| 음식점<br>카페<br>술집  | 한식, 분식, 중식, 양식, 뷔페, 일식 등<br>테마카페, 일반카페, 이색카페, 다방 등<br>호프, 바, 와인바, 칵테일 전문점 등 |
| 문화생활시설<br>레저/스포츠 | 도서관, 전시장, 과학관, 박물관, 극장 등<br>당구장, 골프연습장, 볼링장, 각종경기장 등                        |
| 쇼핑               | 전문매장/상가, 시장, 아울렛, 백화점 등                                                     |

우리는 먼저, 운전자 개인의 취향이 반영되기 쉬운 카테고리에 속하는 POI들만을 추천 대상으로 삼는다. 이를 위해 선별한 카테고리들은 표 1에 나온 바와 같다. 우리는 이들 중 상호 연관성이 깊은 음식점, 카페, 술집을 묶어 음식점/카페/술집의 단일 카테고리로, 문화생활시설과 레저/스포츠를 묶어 문화생활시설/레저/스포츠의 단일 카테고리로 구성한다. 따라서 쇼핑까지, 총 세 가지의 카테고리에 대하여, 카테고리 별 모델 학습 및 개인화 POI 추천을 수행한다.

## 3.1 개인화 POI 추천 모델 학습

본 논문에서 제안하는 (카테고리 별) POI 추천 모델은 SAE-NAD를 기반으로 하며, 구조는 그림 1과 같다. 모델의 입력으로서, 운전자의 POI 방문 횟수, 방문 후 남긴 피드백 (즉, Like/Dislike), POI들 간의 거리 정보가 사용된다. 이러한 입력 벡터를 잘 복원하기 위한 방향으로의 모델 학습이 수행된다. 그 후, 출력으로서는, 운전자가 현재 위치한 지역 내 POI에 대한 해당 운전자의 방문 예측 확률이 도출된다.

구체적으로 살펴보자면, 우리는 먼저 입력으로 들어온 타겟 운전자의 POI 방문 횟수와 해당 운전자가 방문 후 남긴 피드백을 학습 과정에서의 가중치로서 이용한다. 예를 들어, 그림 1에서의 POI A는 운전자의 방문 횟수가 많고 긍정적인 피드백을 받은 POI이다. 따라서, 해당 운전자가 선호하는 POI일 가능성이 더 높도록 높은 가중치와 함께 학습된다. 반면, POI B는 방문 횟수도 적고 부정적인 피드백을 받았으므로, 운전자가 선호하는 POI일 가능성이 더 낮도록 낮은 가중치와 함께 학습된다. 가중치에 대한 계산은 식 1에 나온 바와 같이 수행된다. 식 1에서,  $r_{u,i}$ 는 운전자  $u$ 의 POI  $i$ 에 대한 방문 횟수를 가리킨다.

$$c_{u,i} = \begin{cases} 1 + 2 \log \left( 1 + \frac{r_{u,i}}{e^4} \right) & \text{if } i \text{ has received positive feedback from } u, \\ 1 + 2 \log \left( 1 + \frac{r_{u,i}}{e^6} \right) & \text{if } i \text{ has received negative feedback from } u, \\ 1 & \text{otherwise.} \end{cases} \quad (1)$$

이러한 학습 과정을 통해, 우리는 운전자가 방문한 적 있던 POI들 간에도, 방문 횟수 및 피드백에 따라 차등을 두어 운전자의 개인 취향이 더 잘 추론될 수 있도록 만든다.

그림 1에서의 디코더는 운전자의 POI 방문 히스토리 복원 시, SAE-NAD에서와 같이 운전자가 방문한 적 있던 POI들과 가까운 거리에 위치한 POI일수록 방문 예측 확률이 더 높아지도록 계산한다. 학습이 끝난 디코더에 의해 방문 확률이 높게 예측된

상위 K개의 POI들을, 우리는 해당 운전자의 개인 취향이 잘 반영된 즉, 해당 운전자가 방문할 가능성이 높은 POI들로 간주한다.

### 3.2 운전자 그룹 인기도 반영

개인화 POI 추천 모델에 의해 선별된 타겟 운전자를 위한 상위 K개의 POI들에 대하여, 해당 POI들이 타겟 운전자와 동일한 성별/연령대를 갖는 운전자들로부터 얼마나 많이 방문된 적이 있는지 즉, 인기있는 정도를 확인한다.

이를 위해 우리는 먼저 성별은 남/여, 연령대는 20-30/40-50/60대 이상으로 구분하여 성별/연령대에 따른 총 6개의 그룹을 구성한다. 다음, 타겟 운전자와 동일한 그룹에 속하는 운전자들 중, 개인화 POI 추천 모델에 의해 선별된 K개의 POI 각각을 방문한 적 있는 운전자들의 수를 확인한다. 이를 min-max 정규화 방법을 통해 0-1 사이의 값으로 맞춰준 뒤, 개인화 POI 추천 모델에 의해 도출되었던 방문 예측 확률 값과 곱하여준다.

이러한 곱 연산 결과 기준, 상위 N개의 POI들을 최종적으로 타겟 운전자에게 추천 결과로서 제공한다. 즉, 우리는 이러한 방식을 통해, 운전자의 개인 취향과 잘 부합되면서, 해당 운전자와 동일한 성별/연령대의 운전자들로부터 인기 또한 많은 POI들을 효과적으로 추천할 수 있는 것이다.

## 4. 실험

### 4.1 내비게이션 데이터셋

우리는 본 논문에서, 총 6개월동안 (2021/5-10) 실제 운전자들로부터 수집된 내비게이션 사용 이력 데이터셋을 이용한다. 전처리를 위해, 음식점/카페/술집에 대해서는 8명 미만의 운전자들이 방문한 POI 들과 8개 미만의 방문 POI를 갖는 운전자들을 제거하였고, 문화생활시설/레저스포츠, 쇼핑에 대해서는 3명 미만의 운전자들이 방문한 POI들과 3개 미만의 방문 POI를 갖는 운전자들을 제거하였다. 카테고리 별 데이터셋에 대한 통계적 정보는 표 2에 나온 바와 같다.

표 2. 데이터셋 통계

| 카테고리             | #(운전자)  | #(POI) | #(체크인)    | Density |
|------------------|---------|--------|-----------|---------|
| 음식점/<br>카페/술집    | 141,164 | 65,492 | 2,414,520 | 0.026%  |
| 문화생활시설/<br>레저스포츠 | 100,782 | 19,125 | 771,369   | 0.040%  |
| 쇼핑               | 222,127 | 40,252 | 2,004,133 | 0.022%  |

### 4.2 평가 방식

우리는 운전자의 카테고리 별 POI 방문 히스토리를, 시간 순 기준 8:2로 구분하여 트레이닝셋과 테스트

셋을 나누었다. 이 때, 테스트셋 상에는 운전자가 이전에 방문한 적 있던 POI들까지 모두 포함되어 있다. 운전자의 현재 위치 기준, 동일 광역시/도 내에 위치한 top-N ( $N = 5, 10$ ) 개의 POI들을 선별하여, 비개인화 (인기도 기반의 랭킹 방법) 추천 방법과의 추천 정확도를 비교하였다. 평가를 위한 추천 메트릭으로는 recall (R)을 사용하였다.

### 4.3 실험 결과

표 3. 비개인화 추천 방법과의 추천 정확도 비교

| 카테고리          | 메트릭  | 비개인화  | 개인화          | Gain  |
|---------------|------|-------|--------------|-------|
| 음식점/<br>카페/술집 | R@5  | 0.032 | <b>0.038</b> | 15.7% |
|               | R@10 | 0.054 | <b>0.061</b> | 12.9% |
| 문화/레저         | R@5  | 0.099 | <b>0.102</b> | 3.0%  |
|               | R@10 | 0.152 | <b>0.154</b> | 1.3%  |
| 쇼핑            | R@5  | 0.185 | <b>0.193</b> | 4.3%  |
|               | R@10 | 0.277 | 0.277        | 0%    |

표 3은 각 카테고리에 대하여, 본 논문에서 제안하는 개인화 POI 추천 제안 방법 적용 시 나타나는 효과를 보인다. 우리의 제안 방안은 비개인화 추천 방법 대비, 최대 약 15%의 정확도 향상 결과를 보이며, 이는 지리적 정보와 운전자의 피드백 정보를 사용한 개인화 추천이 매우 효과적임을 입증하는 결과이다.

## 5. 결론

본 연구에서는, 실제 차량 운전자들로부터 수집된 내비게이션 데이터셋을 이용해, 운전자가 방문 후 남긴 리뷰 및 그룹 별 인기를 이용하여 개인화 POI 추천을 수행하는 방법을 새롭게 제안하였다. 실세계 데이터셋을 이용한 실험 수행 결과, 제안 방안이 POI 추천 정확도 측면에서, 매우 효과적임을 확인할 수 있었다.

## 참 고 문 헌

- [1] Ye, M. et al., Exploiting Geographical Influence for Collaborative Point-of-Interest Recommendation, ACM SIGIR 2011.
- [2] Zhao, S. et al., Geo-Teaser: Geo-Temporal Sequential Embedding Rank for Point-of-Interest Recommendation, WWW 2017.
- [3] Wang, H. et al., Exploiting POI-Specific Geographical Influence for Point-of-Interest Recommendation, IJCAI 2018.
- [4] Ma, C. et al., Point-of-Interest Recommendation: Exploiting Self-Attentive Autoencoders with Neighbor-Aware Influence, CIKM 2018.