

증강 그래프 기반 그래프 뉴럴 네트워크를 활용한 POI 추천 모델

정현지¹, 장광선²

¹ 국립공주대학교 인공지능학부 조교수

² 한국과학기술정보연구원 NTIS센터 선임연구원

hjjeong@kongju.ac.kr, gsjang@kisti.re.kr

Next POI Recommendation based on Graph Neural Network of Augmented Graph

Hyun Ji Jeong¹, Gwangseon Jang²

¹School of Artificial Intelligence, Kongju National University

²NTIS Center, Division of National S&T Data, Korea Institute of Science
and Technology Information(KISTI)

요약

본 연구는 궤적 데이터(trajectory data)를 대상으로 증강 그래프 기반의 그래프 뉴럴 네트워크를 활용하여 다음에 방문한 장소를 추천하는 모델을 제안한다. 제안 모델은 전체 궤적 데이터를 그래프로 표현하여 추출한 글로벌 궤적 플로우의 특성을 다음 방문할 POI 추천에 활용한다. 이때, POI 추천 시 자주 발생하는 두 가지 문제를 추가로 해결함으로써 POI 추천의 정확도를 높이는 것을 목표로 한다. 첫 번째 문제는 추천 대상 궤적 데이터의 길이가 짧은 경우에 성능 저하가 발생한다는 것이다. 두 번째 문제는 콜드-스타트 문제이다. 기존 POI 추천 모델은 매우 적은 방문 기록만 가지는 사용자 또는 POI에 대해서는 매우 낮은 예측 성능을 보인다. 본 연구에서는 궤적 그래프에서 일부 옛지를 삭제하여 생성한 증강 그래프 기반의 궤적 플로우 특징 기반 모델을 제안함으로써 짧은 길이의 궤적 데이터 및 콜드-스타트 사용자/POI에 대한 추천 성능을 높인다.

1. 서론

스마트폰의 발전과 위치 기반 소셜 네트워크 서비스(LBSN, location-based social network)의 보급 확대로 인해 LBSN 서비스는 초대용량의 실시간 체크인 데이터를 수집할 수 있게 되었다. 실시간 체크인 데이터는 위치 정보 기반의 다양한 서비스로 활용될 수 있는데, POI 추천이 대표적인 서비스로 주목받고 있다. POI 추천 모델은 사용자가 방문한 장소 궤적 데이터(trajectory data)를 기반으로 다음에 방문할 만한 장소를 추천한다.

POI 추천 모델의 주요 이슈는 일반적으로 짧은 길이의 궤적 데이터에 대해 성능 저하가 발생한다는 것이다[1,2]. 기존의 대부분 모델은 정보량이 많은 긴 길이의 궤적 데이터를 대상으로 다음 방문할 장소를 추천하는 경우는 상대적으로 좋은 성능을 보이지만 짧은 길이의 궤적 데이터를 대상으로 성능이 급격하게 저하된다. 실제 데이터에서는 짧은 길이의 궤적 데이터가 대부분을 차지하기 때문에 이를 해결하는 것은 매우 중요하다. 또한, 추천 모델의 대표적

인 문제인 콜드-스타트 문제가 POI 추천 모델에서도 여전히 발생한다. 매우 적은 양의 체크인 데이터가 존재하는 사용자 또는 POI에 대해서 정확도의 감소가 두드러진다.

본 연구에서는 이를 해결하기 위해 증강 그래프 기반의 그래프 뉴럴 네트워크를 활용한 POI 추천 모델을 제안한다. 먼저, 궤적 플로우 그래프를 생성하는 데 이는 데이터에 존재하는 모든 POI 간의 궤적들을 하나의 그래프로 요약하여 POI의 글로벌한 특징을 추출하고자 함이다. 증강 그래프는 궤적 플로우 그래프에서 일부 옛지를 삭제하여 생성한다. 이는 일부 옛지를 삭제함으로써 짧은 길이의 궤적 데이터 및 매우 적은 양의 체크인 데이터의 형태를 포함하는 그래프를 임의로 생성하는 효과를 지닌다. 따라서, 증강 그래프를 활용하여 훈련한 특징은 짧은 길이의 궤적 데이터 및 콜드-스타트 사용자/POI의 특징을 효과적으로 반영한다.

2. 문제 정의

사용자 집합 $U = \{u_1, u_2, \dots, u_M\}$ 과 POI 집합

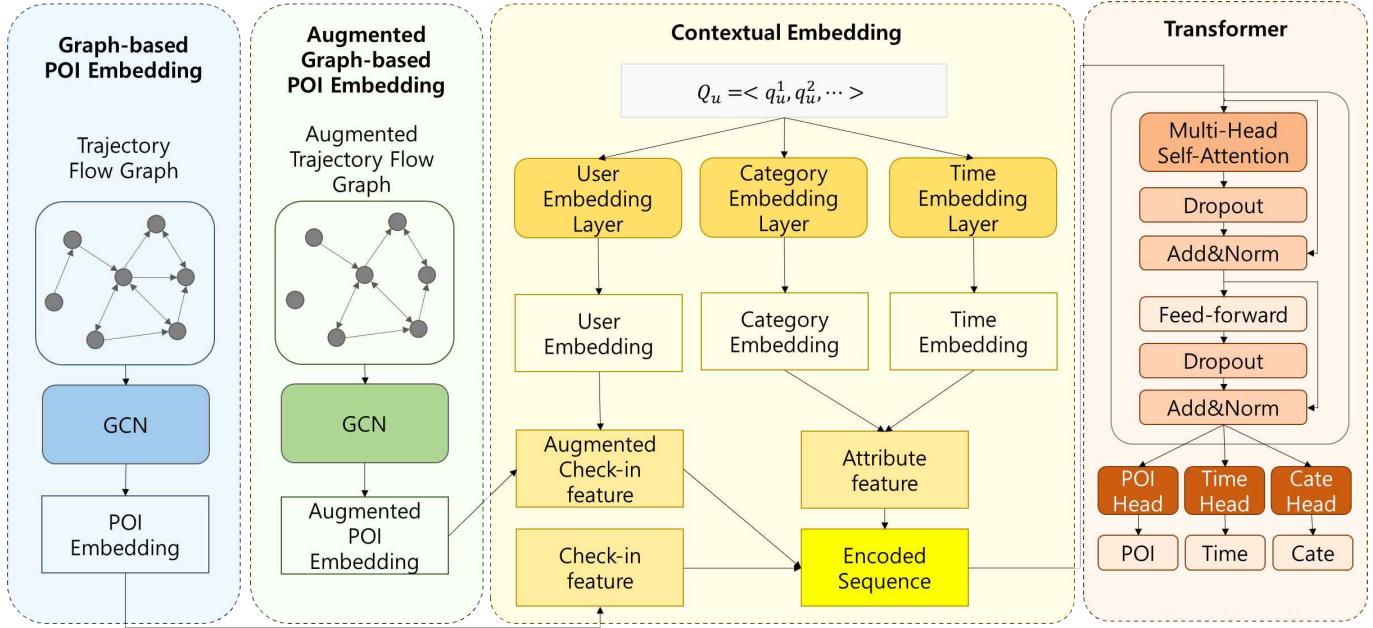


그림 1 증강 그래프 기반 그래프 뉴럴 네트워크를 활용한 POI 추천 모델

$P = \{p_1, p_2, \dots, p_N\}$, 타임스탬프 집합 $T = \{t_1, t_2, \dots, t_K\}$ 가 주어졌다고 가정하자. 이 때, M, N, K 는 양수이다. 각 POI는 위도 lat, 경도 lon, 카테고리 cat, 체크인 빈도 freq의 튜플 $<lat, lon, cat, freq>$ 이다. 체크인은 사용자 u 가 POI p 에 타임스탬프 t 에 방문한 정보를 포함하며 튜플 $q = <u, p, t>$ 로 나타낸다. 사용자 u 의 체크인 쿼리 데이터 $Q_u = (q_u^1, q_u^2, q_u^3, \dots)$ 에서 q_u^i 는 i 번째 체크인 데이터를 나타낸다. POI추천은 사용자 u 와 체크인 쿼리 데이터 Q_u 가 주어졌을 때, 사용자 다음의 POI를 예측하는 것이다.

3. 제안 모델

그림 1은 본 논문에서 제안한 모델을 도식화한 그림이다. 제안한 모델은 총 4가지 모듈로 구성된다.

- 그래프 기반 POI 임베딩(Graph-based POI embedding)
- 증강 그래프 기반 POI 임베딩(Augmented Graph-based POI embedding)
- 콘텍스트 임베딩(Contextual Embedding)
- 트랜스포머(Transformer)

그래프 기반 POI 임베딩을 위해서 먼저 쿼적 플로우 그래프를 생성하는 데 이는 전체 데이터의 POI를 노드로 쿼적에 존재하는 체크인 쌍 (q_u^i, q_u^j) 을

엣지로 하는 그래프이다. POI의 위도, 경도, 카테고리, 빈도는 POI노드의 속성으로 할당하고 각 엣지의 가중치는 체크인 쌍의 발생빈도로 정의한다. 다음으로 쿼적 플로우 그래프에서 GCN을 활용하여 노드인 POI의 임베딩을 추출한다. 증강 그래프 기반 POI 임베딩은 증강 그래프에서 GCN을 하여 POI 임베딩을 수행한다. 증강 그래프는 앞서 생성한 쿼적 플로우 그래프에서 약 20%의 엣지를 랜덤하게 제거하여 생성한다.

콘텍스트 임베딩 모듈은 사용자 임베딩, 카테고리 임베딩, 타임 임베딩으로 구성된다. 사용자 임베딩 벡터와 카테고리 임베딩 벡터는 원-핫 인코딩 벡터를 밀집벡터로 임베딩하여 생성하는데 이들은 추후에 사용자별 쿼적 데이터를 활용하여 훈련된다. 시간 임베딩 벡터는 time2vector[3]를 기반으로 생성한다. 그런 후에 POI 임베딩 벡터와 사용자 임베딩 벡터를 결합하고, 카테고리 임베딩 벡터와 시간 임베딩 벡터를 결합하여 사용자 체크인에 관한 특징과 속성에 대한 특징 벡터를 생성한다. 그런 후에 이를 결합하여 트랜스포머의 입력으로 사용한다. 트랜스포머에서는 앞서 생성한 특징을 입력으로 받아 POI, 시간, 카테고리 정보를 결과로 도출한다. 훈련은 각 POI, 시간, 카테고리의 라벨을 표현한 원-핫 벡터와 트랜스포머에서 도출한 결과 벡터 간의 크로스엔터로피 기반 손실 함수를 사용한다.

4. 실험 결과

본 연구의 성능을 검증하기 위해 Foursquare-NYC 데이터[4]를 활용하여 실험을 수행하였다. Foursquare-NYC 데이터의 통계는 아래 표와 같다.

표 2 Foursquare-NYC 통계

	Foursquare-NYC
사용자	1,075
POI	5,099
카테고리	318
체크인	104,074
궤적	14,160

MAP을 활용하여 정확도를 평가하였으며, 제안 모델은 Foursquare-NYC에 대한 성능은 상위 20개의 결과를 기준으로 MAP을 계산했을 때, 0.3470의 정확도를 보인다.

4. 결론

본 논문은 다음에 방문할 POI를 추천하는 모델을 제안한다. 특히, 증강 그래프에서의 POI 임베딩을 활용하여 짧은 궤적 데이터에서의 성능 저하 및 콜드-스타트 사용자/POI에 대한 추천 성능을 높이는 방법을 고안하였다. 향후 제안한 아이디어를 다양한 데이터에 적용하고 실험을 확대하고자 한다.

감사의 글

이 논문은 2023년도 공주대학교 학술연구지원사업의 연구지원에 의하여 연구되었음.

참고문헌

- [1] Yang, Song, Jiamou Liu, and Kaiqi Zhao. "GETNext: trajectory flow map enhanced transformer for next POI recommendation." Proceedings of the 45th International ACM SIGIR Conference on research and development in information retrieval. 2022.
- [2] Wang, Xinfeng, et al. "EEDN: Enhanced Encoder-Decoder Network with Local and Global Context Learning for POI Recommendation." Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval. 2023.

[3] Kazemi, Seyed Mehran, et al. "Time2vec: Learning a vector representation of time." arXiv preprint arXiv:1907.05321 (2019).

[4] Yang, Dingqi, et al. "Modeling user activity preference by leveraging user spatial temporal characteristics in LBSNs." IEEE Transactions on Systems, Man, and Cybernetics: Systems 45.1 (2014): 129–142.