Geographical POI 요약본

#소개

기존 방식의 추천 방법은 두 지점 간의 지리적 영향력을 물리적 거리를 고려했을 때 동일한 사용자가 두 지점을 함께 방문할 확률 또는 성향으로 모델링합니다. 이러한 방법은 지리적 영향력이 물리적 거리에 의해 결정된다고 가정하기 때문에 지리적 영향력의 비대칭성과 지리적 영향력의 높은 변동성을 포착하지 못합니다

해당 문서는 POI별 지리적 영향력을 활용하여 POI 추천을 개선합니다. 지리적 영향력, 지리적 민감도, 물리적 거리라는 세 가지 요소를 사용하여 두 POI 간의 지리적 영향력을 모델링합니다.

즉, 사용자가 한 지점을 선택했을 때 해당 지점과 다른 지점의 지리적 영향력을 고려하여 추천합니다. 이 문서에서 두 개의 실제 데이터 세트에 대한 실험 결과에 따르면 POI별 지리적 영향력이 POI 추천의 성능을 크게 향상시키는 것으로 나타났습니다.

#용어 정의

1. **지리적 영향력**: 이 용어는 주변의 다른 관심 지점들에게서 관심 지점(POI)이 받는 영향력이나 매력을 나타냅니다. 관심 지점이 다른 장소로부터 방문자를 유인할 가능성을 나타냅니다. 예를 들어, 지하철 역은 다양한 위치에서 사람들을 끌어들이는 교통 허브로 작용하기 때문에 지리적 영향력이 높을 수 있습니다.
2. **지리적 감수성**: 이 용어는 관심 지점이 다른 관심 지점으로부터 방문자를 받을 가능성을 나타냅니다. 관심 지점이 다른 장소로부터 방문자를 받는 정도를 나타냅니다. 예를 들어, 식당은 사람들이 식사나 사회적 활동을 위해 자주 모이는 장소이므로 지리적 감수성이 높을 수 있습니다.

#관련 기술들

POI 추천의 가장 두드러진 특징 중 하나는 POI와 타겟 사용자의 위치가 추천에 중요한 요소라는 점입니다.

지리적 영향력 POI추천 패러다임(2가지)

1. 지리적 근접성을 활용하여 사용자 선호도 학습을 개선하는 방식, 서로 가까운 곳에 있는 POI가 유사한 사용자 선호도를 공유한다고 가정한다. 이러한 방법에서는 지리적 근접성을 사용자 선호도에 대한 일종의 공간 정규화로 사용한다.
2. 물리적 거리를 고려할 때 동일한 사용자가 두 POI를 함께 방문할 확률 또는 성향으로  
   POI간의 지리적 영향력을 명시적으로 수정한다. 물리적 거리에 대한 스펙트럼을 가진 POI의 공동 방문 확률 분포를 포착하기 위해 파워법칙 함수, 가우스 분포등 다양한 형태의 함수가 사용된다.

위의 방법들은 지리적 영향력을 활용하는데 어느정도 성공적이지만, 지리적 영향력의 높은 변동성을 POI간 포착할 수는 없다. 그림 1에서 볼 수 있듯이 Foursquare 데이터 세트에서 무작위로 선택된 10개의 POI는 지리적 영향력이 상당히 다르게 나타나며, 이는 지리적 영향력이 물리적 거리만으로는 잘 포착될 수 없으므로 지리적 영향력을 POI별로 파악해야 한다는 것을 나타냅니다.

이 문서에서는 POI 추천을 개선하기 위해 POI별 지리적 영향력을 활용합니다. 지리적 영향력, 지리적 민감도, 물리적 거리라는 세 가지 요소를 사용하여 두 POI 간의 지리적 영향력을 모델링합니다. 예를 들어, 지하철 역은 일반적으로 지리적 영향력이 높고 레스토랑은 일반적으로 지리적 민감도가 높습니다. 여기서 지리적 영향력과 지리적 민감도는 두 개의 저차원 벡터이며, 두 POI 간의 지리적 영향력은 한 POI의 지리적 영향력 벡터와 다른 POI의 지리적 민감도 벡터의 내적 곱으로 표현됩니다

# 지리적 영향력을 고려한 추천모델 장점

POI별 지리적 영향력 모델에는 두 가지 고유한 이점이 있습니다: (1) POI 간의 지리적 영향력은 비대칭적이기 때문에 POI 간 지리적 영향력의 높은 가변성을 포착할 수 있는 높은 유연성을 제공합니다. 즉, 지하철이라는 POI 지점은 지리적 영향력은 높지만 지리적 감수성은 낮습니다. 사람을 끌어모은다는 점에서 지리적 영향력은 높지만 사람들이 즐기기 위해 가는 장소가 아니므로 지리적 감수성은 낮습니다.

(2) POI 상호작용 행렬을 사용하여 POI별 지리적 영향력을 직접 모델링하는 대신, 우리 모델은 각 POI에 대해 두 개의 저차원 벡터로 지리적 영향력을 표현하여 자유 파라미터의 수를 크게 줄입니다[Wang et al., 2015b]. 따라서 데이터 희소성 문제가 심각한 POI 추천에 적합한 모델입니다.

#지리적 영향력을 고려한 추천모델의 성능

현 문서의 추천 모델은 사용자의 선호도를 파악하는 표준 모델에 POI별 지리적 영향력을 통합하여 새로운 POI 추천 방식을 구축합니다. 사용자의 체크인 기록을 사용하여 모델을 학습시키고, 사용자가 가까운 미래에 방문할 가능성이 높은 POI를 '예측'하는 데 모델을 적용하여 추천 성능을 검증합니다. 포스퀘어와 고왈라의 두 가지 실제 데이터 세트에서 광범위한 실험을 수행하여 모델의 효율성을 입증했습니다. 실험 결과, POI별 지리적 영향력이 POI 추천의 성능을 크게 향상시켜 최신 POI 추천 방법을 능가하는 것으로 나타났습니다.

#기존의 POI 추천 모델

POI 추천모델은 사용자의 체크인 기록에 따라 사용자가 방문하지 않은 POI를 추천하는 기능입니다. 사용자의 체크인은 암묵적 피드백이라는 점을 고려할 때, 기존 방식은 체크인 횟수에서 변환된 점수로 피팅하여 체크인을 모델링하거나[Lian et al., 2014], 사용자의 선호도 순위를 POI에 최적화하여 모델링합니다

최종적으로 다음과 같은 POI 추천시스템

POI 추천: 체크인 기록이 있는 사용자 집합 U와 위치 정보(경도, 위도)가 있는 POI 집합 I가 주어지면, POI 추천은 각 대상 사용자 u ∈ U에 대해 타켓 사용자가 잠재적으로 관심이 있지만 추천하기전까지 방문하지 않은 POI로 구성된 {i|i ∈ I}의 목록을 추천합니다.

#지리적 영향력을 고려한 추천모델 설계

POI별 지리적 영향력 모델 설계를 하기 위해서 논문에서는 다음과 같이 지리적 영향력을

정의한다. 지리적 영향력 -> yij = ~gT i ~hj × f(dij ) ,

y는 POI i에 대한 POI j의 지리적 영향력, 즉 y는 i가 j를 추천하게 되는 기준을 의미한다.

숙박시설(호텔)이라는 POI i가 있다면 목표 POI j(놀거리, 먹거리)에 대해 지리적 영향력을

고려하여 사용자에게 POI j를 추천합니다. 지리적 영향력이 높을수록 추천가능성이 높아지고

지리적 영향력이 낮을수록 추천가능성이 낮아집니다.

지리적 영향력을 고려할 때 yij = ~gT i ~hj × f(dij ) 해당 식의 요소를 살펴보자.

g는 POI i(호텔)지점의 지리적 영향력이다. 즉, 해당 지점이 얼마나 많은 사람들을 끌어 들이냐? 인데 이를 해당 논문에서는 사용자의 체크인 이력으로 값을 정의한다.

h는 POI j(놀거리, 먹거리)의 지리적 감수성이다. 즉, 해당 지점이 다른지점으로부터 사용자를 받는 정도를 나타낸다. 이를 해당 논문에서는 어떤걸로 정의한다고 명시되어 있지 않다.

~gT i ~hj 해당 식의 값은 두 저차원 벡터를 내적한 값이다. 즉, POI i의 지리적 영향력 벡터와

POI j에 지리적 감수성 벡터를 내적한 값이다.

f(dij )는 두 POI의 물리적 거리 dij 가 주어졌을 때 동일한 사용자가 두 POI를 방문할 확률을 반영합니다. 이 문서에서는 Table 2에 표시된 네 가지 유형의 함수를 고려합니다. 일반적으로 f(dij )는 dij 가 증가함에 따라 감소하며, 이는 사용자가 지리적으로 인접한 POI를 방문하는 것을 선호하는 현상을 포착합니다. 또한, f(dij )의 파라미터는 POI 추천 모델을 학습하기 전에 미리 학습하거나 POI 추천 모델 학습과 함께 학습할 수 있습니다.

사용자 u가 방문한 POI의 집합 Hu와 “지리적 영향력”이 주어지면, 모든 방문 POI의 영향을 고려하고 대상 POI j에 대한 Hu의 전반적인 지리적 영향을 (2)와 같이 모델링합니다.



위의 식을 해석해보자면 POI j가 사용자가 방문한 지점 Hu에 대하여 두 POI 간의 지리적 영향력을 계산하는 모델이다.

이번에는 타겟 POI에 대한 사용자의 선호도를 추론하기 위해 사용자 선호도와 지리적 영향력의 영향을 모두 고려합니다. 사용자 u의 POI j에 대한 선호도를 suj라고 정의합니다.



여기서 ~tu와 ~zj는 사용자 u의 선호도와 POI j의 선호도 간의 상호작용을 모델링하는 데 사용되는 행렬 인수 분해 방법에 따라 주어집니다.

즉, 해당식을 쉽게 풀이하자면 “사용자 체크인 기반 추천모델” + “지리적 영향력 모델”로

사용자 u가 POI j에 대해서 지리적 영향력을 고려한 선호도 모델을 정의하였다.

Suj를 이용하여 사용자 u가 POI j를 선호할 확률을 Puj로 정의합니다.



여기서 I는 POI의 집합이고 분모는 주어진 사용자 u에 대한 모든 POI의 정규화입니다

e^n을 뜻하는 exp(Suj)를 통하여 전체 POI j 중에서 사용자가 해당 POI j를 좋아할 확률 구하는 Puj를 구한다.

즉, 사용자 u가 관심지점 POI j에 대해 각각의 선호도 값의 합을 분모로 만들고

사용자가 실제적으로 선택한 관심지점 POI j를 분자로 만들어 확률을 구한다.

마지막으로 사용자 u가 주어졌을 때, 사용자가 POI j를 방문할 확률 puj에 따라 미방문 POI를 추천합니다. 각 사용자의 추천 목록은 미방문 POI 중 확률 puj가 가장 높은 상위 n개의 POI로 구성됩니다.

#논문에서의 아이디어

해당 논문을 읽고 다음과 같은 아이디어를 생각했다.

지리적 영향력을 고려할 때 yij = ~gT i ~hj × f(dij )를 사용하는데

POI i에 대한 지리적 영향력에 대해서는 해당 관광지의 지하철, 버스등과 같은

교통수단의 수로 가중치를 부여한다.

POI j에 대한 지리적 감수성에 대해서는 POI j에 대한 카테고리를 음식점으로 축소하고

평점으로 가중치를 부여한다.

두 지점사이의 물리적 거리를 의미하는 f는 유지한다.

해당 식을 이용하여 지리적 영향력을 결정하고 이를 MF모델에 결합하여

지리적 영향력을 고려한 개인화 POI에 대한 아이디어를 도출했다.