

협업 관심 지점 추천을 위한 지리적 영향력 활용하기

-요점

본 논문에서는 관심 지점(POI) 추천 서비스를 제공하는 것을 목표로 합니다.

추천 서비스를 제공하는 것을 목표로 합니다.

Foursquare, Whrrl 등 빠르게 성장하고 있는 위치 기반 소셜 네트워크(LBSN-Location-based Social Net works)에 대해 설명합니다.

저희의 아이디어는 사용자 선호도, 사회적 영향력, 지리적 영향력 등을 고려하여 POI를 추천하는 것입니다. 사용자 기반 협업 필터링을 기반으로 사용자 선호도를 도출하고 친구의 사회적 영향력을 탐색하는 것 외에도, 사용자의 체크인 활동에서 나타나는 공간적 군집화 현상에 따른 지리적 영향력에 특히 주목했습니다. 우리는 사용자 체크인 행동에 지리적 영향력이 중요한 역할을 한다 주장하고, 이를 파워 법칙 분포로 모델링합니다.

이에 따라 지리적 영향력에 기반한 협업 추천 알고리즘을 naïve Bayesian(나이브 베이지안) 기반으로 개발합니다. 또한 사용자 선호도와 사회적 영향력, 지리적 영향력을 결합한 통합 POI 추천 프레임워크를 제안합니다.

마지막으로 Foursquare와 Whrrl에서 수집한 두 개의 대규모 데이터 세트에 대한 종합적인 성능 평가를 수행합니다. 이러한 실제 데이터 세트에 대한 실험 결과에 따르면 통합 협업 추천 접근 방식이 광범위한 다양한 추천 접근 방식보다 훨씬 뛰어난 성능을 보여줍니다.

1. 소개

모바일 기기, 무선 네트워크 및 웹 2.0 기술의 급속한 발전으로 최근 몇 년 동안 Loopt1, Brightkite2, Foursquare3, Whrrl4와 같은 위치 기반 소셜 네트워킹 서비스가 다수 등장했습니다.

이러한 LBSN을 통해 사용자는 친구나 다른 사용자와 사이버 링크를 설정하고 레스토랑, 상점, 영화관 등 다양한 관심 지점(POI)에 대한 방문 팁과 경험을 공유할 수 있습니다.

LBSN에서, **POI 추천 서비스**-(새로운 장소를 탐색하고 도시를 더 잘 알 수 있도록 사용자에게 새로운 POI를 추천하는 것을 목표로 함.)는 최근 많은 연구가 진행되고 있는 필수 기능입니다. [25, 26]

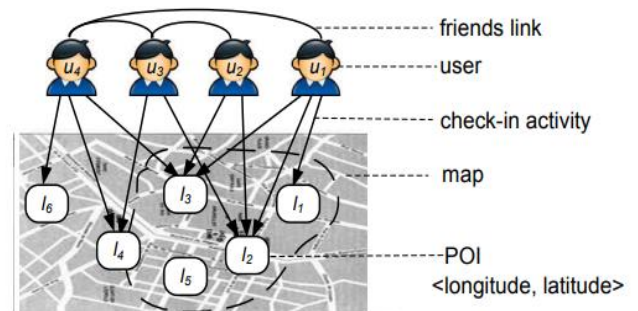


그림 1: LBSN에서, 사용자 간 친구관계 및 사용자 위치 체크인 활동의 그래프 표현

실제로, LBSN에서 POI 추천을 가능하게 하는 것은 유망하고 흥미로운 연구 문제입니다.

사용자 간의 '사이버' 연결뿐만 아니라 사용자와 위치 간의 '물리적' 상호 작용과 같은 귀중한 정보가 시스템에 캡처되어 있기 때문입니다.

그럼에도 불구하고, 이러한 정보는 POI 추천과 관련된 선행 연구에서 충분히 탐구되지 않았습니다.

예를 들어, Zheng 등은 위치 추천을 위해 모바일 사용자의 GPS 궤적 로그에서 방문한 위치를 추출했습니다 [25, 26].

1. www.loopt.com
2. www.brightkite.com
3. www.foursquare.com
4. www.whrrl.com

이러한 서비스는 종종 위치 기반 소셜 네트워크라고 불리며, 따라서 논문에서는 LBSN으로 축약됩니다. 그러나 이 연구들은 추천 과정에서

사용자 간의 소셜 연결이나 사용자와 위치 간의 상호작용을 고려하지 않았습니다

본 논문에서는 이 지역의 고유한 지리적 의미를 활용하는 것을 목표로 합니다. LBSN의 POI 추천을 위해 사용자친구의 사회적 영향력을 적용하는 것 외에도 사용자와 위치의 상호작용을 기반으로 합니다. 사용자와 POI는 LBSN에서 필수적인 두 가지 유형의 엔티티입니다. 그림 1에서 볼 수 있듯이, u_1, u_2, u_3, u_4 로 표시되는 LBSN의 사용자는 소셜 링크를 통해 서로 연결되어 사용자 소셜 네트워크를 형성합니다. 또한, l_1, l_2, \dots, l_6 로 표시되는 POI는 일반적으로 다양한 POI에 대한 사용자의 취향을 반영하는 '체크인' 활동을 통해 사용자와 연결됩니다. 마지막으로, 그림에서 논리적으로 설명한 것처럼 (경도, 위도)로 지오코딩된 POI는 지리적으로 제약이 있습니다. 사용자에게 POI를 추천하려면 분명히 이전 사용자의 체크인 활동 기록이 매우 유용합니다. 이러한 정보가 위치 기반 서비스 네트워크(LBSN)에서 제공되므로, 많은 성공적인 CF 기반 추천 시스템에서 POI를 'item'으로 취급하여 기존의 협업 필터링(CF) 기법을 사용하는 것이 POI 추천을 지원하기 위한 직관적인 아이디어가 될 수 있습니다. 이 아이디어의 기본 논거는 이전 체크인 활동에서 POIs와 유사한 방문 행동을 보인 다른 사용자를 통해 사용자의 취향을 추론할 수 있다는 것입니다. 따라서 사용자 기반 또는 항목 기반 협업 필터링 기술을 POI 추천에 적용할 수 있습니다. 또한, LBSN에서 손쉽게 이용할 수 있는 사용자들의 소셜 네트워크를 탐색하여 POI 추천의 성능을 향상시킬 수 있습니다. 최근 연구에 따르면 소셜 친구는 공통 관심사를 공유하는 경향이 있으므로 다음과 같은 프로세스에 사용할 수 있다고 합니다.

추천을 위한 협업 필터링 [11, 12, 2, 13, 24].

위의 아이디어는 LBSN에서 사용할 수 있는 필수 정보, 즉 사용자-위치 상호작용 및 사용자-사용자 소셜 링크를 탐색하는 것을 목표로 하지만, 우리는 사용자의 활동과 지리적 근접성에 자연스럽게 존재하는 지리적 영향을 무시할 수 없다고 주장합니다. 토블러(Tobler)의 지리학 제1법칙에 따르면 "모든 것은 다른 것과 관련되어 있지만 가까운 것들은 먼 것보다 더 관련이 있다"[19]. 따라서 사용자 직관적으로 가까운 POI를 방문하는 경향이 있다. POI 추천에 대한 이러한 직관에서 파생될 수 있는 두 가지 주요 의미는 다음과 같습니다.

(1) 사람들은 집이나 사무실에 가까운 POI를 방문하는 경향이 있습니다.

(2) 사람들은 집에서 멀리 떨어져 있더라도 자신이 선호하는 POI의 인근 POI를 탐색하는데 관심이 있을 수 있습니다.

예: (사용자가 브로드웨이 공연을 보러 타임스퀘어에 갈 때 타임스퀘어 주변의 레스토랑과 상점을 탐색할 수 있음)
LBSN의 지리적 특성으로 인해 우리는 사용자와 POI 사이의 지리적 영향은 물론 POI 간의 지리적 영향도 사용자 사이의 사회적 영향력만큼 중요하다고 믿습니다. 이는 앞서 언급한 바와 같이 LBSN에서 POI 권장 사항을 지원하는 데 긍정적인 역할을 할 수 있습니다. 간단히 말해서, 우리는 LBSN의 POI 추천에 대한 지리적 영향과 사회적 영향을 연구하는 데 관심이 있습니다

LBSN에서 POI 추천 서비스를 지원하기 위한 우리의 접근 방식은 세 가지 보완 요소를 통합하여 사용자의 관심 POI를 발견하는 효과적인 협업 추천 기술을 개발하는 것입니다. i) POI의 사용자 선호도; ii) 사회적 영향력; iii) 지리적 영향. POI에 대한 사용자의 암시적 선호도는 POI에 대한 체크인 활동에서

파생될 수 있습니다. 공통 POI에 많이 체크인한 두 명의 사용자를 유사 사용자로 간주함으로써 유사 사용자의 이전 체크인 활동을 통해 사용자의 암묵적 선호도를 발견할 수 있습니다. 그림 1의 예를 떠올려 보세요. u1, u2 자주 방문하는 많은 POI를 공유하므로, 유사한 체크인 행동, 즉 POI 선호도를 공유하는 것으로 간주되는 유사한 사용자로 간주됩니다. 결과적으로, I1 장소는 u2에게 추천하기 좋은 후보입니다. 이전에 u1이 POI(I1)를 방문했기 때문에. 즉 POI 선호도를 공유하는 것으로 간주되는 유사한 사용자로 간주됩니다. 결과적으로, I1 사용자에게 추천하기에 좋은 후보입니다. u2~부터 u1 이전에 이 POI를 방문한 적이 있습니다. 반면, 친구의 사회적 영향력은 추천 과정에 포함될 수 있습니다. 예를 들어, 고려할 때 I4 장소 추천 후보-> u1, 사회적 영향력: u4->u1 의사결정에 기여할 수 있습니다. 마지막으로, 인근 POI에 대한 POI의 지리적 영향을 고려할 수 있습니다. 예시에서 보듯, u2가 (I2, I3)를 이전에 방문했다. 거기 근처 POI I1, I5 지리적인 영향으로 인해 긍정적으로 여겨질 수도 있다.

앞서 논의한 바와 같이, 사용자 선호도[9, 18, 14, 15, 13]와 사회적 영향력[12, 13]을 기반으로 한 기존의 항목 추천 기법이 POI 추천에 적용될 수 있을 것으로 보인다. 그럼에도 불구하고 LBSN의 POI 권장 사항에 대한 효과는 연구되지 않았습니다. 가장 중요한 것은 POI 추천에 신선하고 유망한 POI 간의 지리적 영향을 통합하려는 아이디어가 이전에 조사된 적이 없다는 것입니다. 본 논문에서는 LBSN에서 사용자 체크인 활동의 "지리적 클러스터링 현상"을 조사하고 POI 간의 지리적 영향을 포착하기 위한 거듭제곱 법칙 확률 모델을 제안합니다. 이에 베이지안 이론을 통해 POI의 지리적 영향력을 통합하여 LBSN을 위

한 타겟 협업 POI 추천 서비스를 구현한다. 마지막으로, 사용자 선호도를 POI에 융합하고, 사용자 간 사회적 영향력과 POI 간 지리적 영향력을 융합하는 통합 위치 추천 프레임워크를 제안한다.

요약하면, 이 연구 작업에서 우리가 이룩한 기여는 네 가지입니다.

- 우리는 위치 기반 소셜 네트워킹 시스템(LBSN)에서 POI 추천 지원 문제를 연구합니다. 여기서 POI는 사용자와 POI 사이에 "물리적" 상호 작용이 필요하다는 사실 때문에 기존 추천 시스템의 다른 추천 item과 고유하게 다릅니다. 따라서 우리는 POI 간의 지리적 영향을 조사하고 LBSN의 POI 추천을 위해 개발하는 협업 추천 기술에 사용자 선호도 및 사회적 영향과 함께 지리적 영향을 통합할 것을 제안합니다.

- 우리는 LBSN에서 사용자 체크인 활동의 지리적 클러스터링 현상을 조사하여 지리적 영향을 포착하기 위한 새로운 아이디어를 개발합니다. 우리는 POI 간의 지리적 영향력을 포착하기 위해 거듭제곱 법칙 확률 모델을 사용하고 naïve Bayesian 방법을 통해 지리적 영향에 기반한 공동 POI 권장 사항을 실현할 것을 제안합니다.

본 논문에서는 사용자 선호도, 사회적 영향력, 지리적 영향력을 융합하여 특정 사용자가 POI를 방문할 확률 예측 모델을 고안함으로써 POI 추천을 위한 통합 추천 프레임워크를 제안한다.

- 마지막으로, 우리는 두 개의 잘 알려진 LBSN(예: Foursquare 및 Whrrl)에서 수집한 대규모 데이터 세트에 대해 제안된 추천 기술

을 평가합니다. 실험 결과는 우리가 제안한 협업 추천 기법이 다른 접근 방식에 비해 우수한 POI 추천 성능을 보인다는 것을 보여줍니다. 우리 평가의 중요한 결과는 아래 요약되어 있습니다.

- 지리적 영향은 사회적 영향보다 LBSN에서 POI 추천의 효과에 더 큰 영향을 미쳐 사회적 영향에 비해 추천 성능을 13.8% 이상 향상시킵니다.
- Random Walk and Restart [12]는 LBSN의 POI 추천에 적합하지 않을 수 있습니다. 실제 데이터와 실험 결과를 분석한 결과, 친구들은 여전히 서로 다른 선호도를 반영하고 있으며, 사회적 유대감은 사용자 간의 체크인 행동의 유사성을 반영하지 못하는 것으로 나타났습니다.
- item 기반 협업 필터링은 현재 LBSN 상태에서 많은 POI가 소수의 사용자 체크인만 표시하기 때문에 사용자 기반 협업 필터링과 비교할 수 있는 결과를 제공하지 않습니다. 따라서 item 유사성은 사용자 유사성만큼 정확하지 않습니다.

이 논문의 나머지 부분은 다음과 같이 구성됩니다. 2장에서는 사용자 선호도와 사회적 영향력에 따른 기존 추천 기법에 대한 배경지식을 제시하고 관련 문헌을 검토한다. 3장에서는 지리적 영향에 따른 위치 추천 과정을 설명한다. 섹션 4에서는 세 가지 요소를 모두 통합하는 위치 추천 프레임워크를 제안합니다. 섹션 5에서는 Foursquare와 Whrrl에서 각각 크롤링된 두 개의 대규모 데이터 세트를 기반으로 다양한 위치 추천 알고리즘에 대한 실증적 연구를 수행합니다. 마지막으로 6장에서 논문을 마무리한다.

2. 예선

이 섹션에서는 먼저 사용자 선호도와 사회적 영향력을 활용하기 위한 융합 접근 방식의 구성

요소 역할을 하는 사용자 기반 협업 필터링 및 친구 기반 추천에 대한 배경을 제공합니다. 다음으로 추천 시스템에 관한 몇 가지 관련 연구를 검토합니다.

2.1 사용자 기반 협업 필터링

협업 필터링을 기반으로 유사한 사용자의 행동을 집계하여 사용자의 암묵적 선호도를 발견할 수 있습니다. U 와 L 은 시스템에서 체크인 활동을 추적하는 LBSN에 설정된 사용자 세트와 POI 세트를 나타냅니다.

사용자 $u_i \in U$ 가 POI $l_j \in L$ 에서 한 체크인 활동은 $c_{i,j}$ 로 표시되며, 여기서 $c_{i,j} = 1$ 은 u_i 가 이전에 l_j 에서 체크인한 적이 있음을 의미하고 $c_{i,j} = 0$ 은 u_i 가 l_j 를 방문한 기록이 없음을 나타냅니다. 이렇게 기록된 사용자 체크인 활동은 사용자의 암묵적인 선호도를 파악하는데 사용되며, 이는 사용자가 방문하지 않은 POI에서 체크인할 가능성을 예측하는 확률로 표현할 수 있습니다. 이 예측을 $\hat{c}_{i,j}$ 로 표시하고 다음과 같이 u_i 에서 l_j 까지의 예측된 체크인 확률을 구합니다.

$$\hat{c}_{i,j} = \frac{\sum_{u_k} w_{i,k} \cdot c_{k,j}}{\sum_{u_k} w_{i,k}} \quad (1)$$

여기서 $w_{i,k}$ 는 사용자 u_i 와 u_k 간의 유사성 가중치입니다. 사용자 u_i 와 u_k 간의 유사도 가중치 $w_{i,k}$ 를 계산하기 위해 코사인 유사도 및 피어슨 상관관계 등 여러 유사도 측정값을 사용할 수 있습니다. 본 연구에서는 단순성 때문에 코사인 유사도를 선택했습니다. 사용자 u_i 와 u_k 사이의 코사인 유사도 가중치는 $w_{i,k}$ 로 표시되며 다음과 같이 정의됩니다.

$$w_{i,k} = \frac{\sum_{l_j \in L} c_{i,j} c_{k,j}}{\sqrt{\sum_{l_j \in L} c_{i,j}^2} \sqrt{\sum_{l_j \in L} c_{k,j}^2}} \quad (2)$$

2.2 친구 기반 협업 필터링

친구들은 친구이고 많은 공통 관심사를 공유할 수 있기 때문에 유사한 행동을 하는 경향이 있으므로 상호 연관된 체크인 행동으로 이어집니다[15, 13]. 예를 들어, 두 명의 친구가 가끔 함께 영화를 보러 놀 수도 있고, 사용자가 친구가 추천하는 레스토랑에 갈 수도 있습니다. 이러한 모든 가능한 이유는 친구가 잠재적인 상관 체크인 행동으로 인해 특정 사용자에게 좋은 추천을 제공할 수 있음을 시사합니다. 즉, 사용자의 친구에게 추천을 요청할 수 있으며 이를 친구의 추천 기반 사회적 영향력이라고 부릅니다. 사회적 영향력에 기반한 POI 추천은 [13]에서 설명한 친구 기반 협업 필터링 접근 방식으로 실현될 수 있습니다

$$\hat{c}_{i,j} = \frac{\sum_{u_k \in F_i} SI_{k,i} \cdot c_{k,j}}{\sum_{u_k \in F_i} SI_{k,i}} \quad (3)$$

여기서 $\hat{c}_{i,j}$ 는 i 에서 j 의 예상 체크인 확률, F_i 는 u_i 의 친구 세트, $SI(k,i)$ 는 u_i 에 대한 방향성 사회적 영향력 가중치입니다[14, 15, 13]. 한편으로는 사회적 유대관계가 긴밀한 친구일수록 추천 측면에서 더 나은 신뢰를 가질 수 있다고 생각합니다. 반면, 더 유사한 체크인 행동을 보이는 친구는 활성 사용자와 더 유사한 취향을 가져야 하므로 해당 친구의 제안이 더 가치가 있습니다. 따라서 다음에서는 위의 두 가지 측면을 결합하여 사회적 영향력 가중치를 도출하는 방법을 소개한다. 두 친구 사이의 사회적 영향력 가중치 도출하는 한 가지 방법은 그들의 사회적 연결과 체크인 활동의 유사성을 기반으로 합니다[12]

$$SI_{k,i} = \eta \cdot \frac{|F_k \cap F_i|}{|F_k \cup F_i|} + (1 - \eta) \cdot \frac{|L_k \cap L_i|}{|L_k \cup L_i|} \quad (4)$$

여기서 η 는 $[0, 1]$ 범위의 튜닝 파라미터이며, F_k 와 L_k 는 각각 사용자 u_k 의 친구 세트와 POI 세트를 나타냅니다.

또 다른 측정 방법은 RWR(Random Walk with Restart) 기술[20]을 사용하여 사용자 간의 사회적 연결과 사용자와 POI 간의 체크인 활동을 모두 캡처하는 그래프를 사용하는 것입니다[24]. 노드 k 에서 시작하여 각 단계마다 무작위로 다른 노드로 연결되는 링크를 따라가며 RWR을 수행합니다. 모든 단계에서 노드 k 에서 다시 시작될 확률이 α 라는 점에 유의하세요. 전체 프로세스가 수렴할 때까지 RWR을 반복적으로 반복하면 각 노드의 정지(또는 정상 상태) 확률을 얻을 수 있습니다. 노드의 고정 확률은 특정 시작 노드(예: 사용자 u_k)에 대한 편향이 주어졌을 때 각 사용자 노드(예: 사용자 u_i)의 장기방문율을 제공합니다. 이는 사용자 u_k 가 u_i 에 미치는 사회적 영향력 가중치, 즉 $SI(k,i)$ 로 해석할 수 있습니다.

2.3 관련 업무

콘텐츠 기반 및 협업 필터링 기술은 추천 시스템에 널리 채택되는 두 가지 접근 방식입니다[1]. 콘텐츠 기반 시스템은 항목 콘텐츠(예: 항목을 설명하는 키워드/태그)와 사용자 프로필 간의 유사성을 기반으로 추천할 항목을 선택합니다[3, 8, 17]

사용자 프로필에 사용된 용어와 항목 콘텐츠 간의 사전 바인딩 관계에 주로 의존하기 때문에 사용자 간의 암시적 연관성은 고려되지 않습니다. 협업 필터링 시스템은 메모리 기반과 모델 기반의 두 가지 범주로 나뉩니다. 메모리 기반 시스템은 다음과 같이 더 분류될 수 있습니다. 사용자 기반 그리고 아이템 기반 시스템. 사용자 기반 시스템의 경우 [9], 모든 사용자 쌍 간의 유사성은 코사인 유사성 또는 Pearson 상관 관계와 같은 일부 선택된 유사성 측정을 사용하여 관련 항목에 대한 평가를 기반으로 계산됩니다. 사용자 유사성을 기반으로, 주어진 사용자-항목 쌍에 해당하는 누락된 등급은 유사한 사용자의 동일한 항목에 대한

등급의 가중치 조합을 계산하여 파생될 수 있습니다. 항목 기반 시스템의 경우 [18], 누락된 등급을 예측하기 위해 사용자 간의 유사성을 사용하는 대신 유사한 항목에 대한 사용자 등급의 가중치 조합을 계산하기 위해 유사한 등급의 항목을 먼저 찾아 예측을 수행합니다. 반면, 모델 기반 협업 필터링 시스템은 사용자가 항목 평가에서 유사한 행동을 기반으로 클러스터를 형성할 수 있다고 가정합니다. 클러스터링 알고리즘이나 베이지안 네트워크와 같은 기계 학습 기술을 사용하여 사용자의 평가 행동에서 인식된 패턴을 기반으로 모델을 학습할 수 있습니다[5, 23].

소셜 네트워킹 시스템의 맥락에서 사회적 우정은 메모리 기반[11, 12] 및 랜덤 워크 기반[2, 11, 12]과 같은 협업 필터링 기반 추천 시스템에 유익한 것으로 나타났습니다. 이들 작품에서는 사회적 친구들은 공통의 관심사를 공유하는 경향이 있으므로 협업 필터링 과정에서 그들의 관계가 고려되어야 한다고 주장합니다. Random Walk는 소셜 네트워크를 확률론적 가중치 링크가 포함된 그래프로 캡처하여 사회적 관계를 나타내므로 항목에 대한 사용자 선호도[12]와 다른 사용자에게 대한 사회적 영향[24]을 정확하게 예측할 수 있습니다. 한편, 모델 기반 시스템에서도 사회적 우정이 탐구되었다 [14, 13]. 이러한 작업은 주로 영화와 같은 항목을 추천하기 위한 기존 추천 시스템에 중점을 둡니다

최근 위치 추천과 마이닝이 연구계로부터 많은 주목을 받고 있다 [27, 26, 7, 25, 22]. 그 중 [27, 26, 7, 25]은 주로 사용자 간의 사회적 관계를 고려하지 않은 GPS 데이터 세트에 중점을 둡니다. 불행하게도 이들 연구에서는 POI 간의 지리적 영향이 탐구되지 않았다[27, 26,

25]. 최근에는 GPS 궤적의 위치 상관관계가 연구되었습니다[7]. 그러나 이 작품에서 장소는 여전히 관습적인 항목으로 취급된다. 이처럼 위치 간의 상관관계는 지리적 영향이 아닌 사용자의 활동을 통해 형성됩니다. [22]는 LBSN에서 위치 추천 서비스를 제공하는 최초의 연구이지만 위치 추천의 효율성을 높이는 것이 목표이다.

우리의 연구는 네 가지 측면에서 이전 연구와 차별화됩니다. i) 캡처된 데이터의 사회적 및 지리적 특징을 모두 포괄하는 위치 기반 소셜 네트워킹 시스템의 적용 영역은 새롭고 독특합니다. ii) LBSN 추천 시스템의 사회적 영향과 지리적 영향에 대한 연구는 이전에 탐구되지 않았습니다. iii) 사용자 선호도, 사회적 영향력과 함께 지리적 영향력을 통합한 통합 협업 추천 접근 방식의 제안은 새롭고 혁신적입니다. iv) 잘 알려진 LBSN에서 수집된 두 개의 대규모 실제 데이터세트, 즉 Foursquare 및 Whrrl 데이터세트를 성능 평가에 채택합니다.

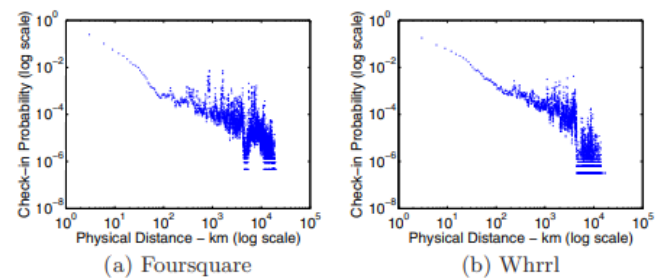


그림 2: 지리적 영향 확률 분포

앞서 언급한 바와 같이 LBSN에서 사용자의 체크인 활동은 POI에서의 물리적 상호작용(즉, 방문)을 기록하므로 POI의 지리적 근접성이 사용자의 체크인 행동에 중요한 영향을 미친다고 주장합니다. 사용자에게 대한 이러한 지리적 영향을 더 잘 이해하기 위해 우리는 잘 알려진 두 개의 LBSN(예: Foursquare 및 Whrrl)에서 수집된 사용자 체크인 활동의 실제 데이터 세트에 대한 공간 분석을 수행합니다. 구체적으로, 우리는 다음을

연구하는 것을 목표로 합니다. 거리사용자의 체크인 POI 중 두 개가 주어진 거리 내에 있을 가능성을 측정하여 사용자 체크인 행동에 대해 알아봅니다. 이 측정값을 얻기 위해 우리는 사용자가 체크인한 모든 POI 쌍 사이의 거리를 계산하고 동일한 사용자가 체크인한 POI의 거리에 대한 히스토그램(실제로 확률 밀도 함수)을 그립니다. 그림 2에서 볼 수 있듯이, 동일한 사용자가 체크인한 POI 쌍의 상당 부분이 근거리 내에 있는 것으로 나타납니다. 지리적 클러스터링 현상사용자 체크인 활동에서 이러한 현상은 다음과 같은 경향으로 직관적으로 설명될 수 있는 지리적 영향에 기인할 수 있습니다.

(1) 사람들은 집이나 사무실과 가까운 POI를 방문하는 경향이 있습니다.

(2) 사람들은 집에서 멀리 떨어져 있더라도 자신이 선호하는 POI의 인근 POI를 탐색하는데 관심이 있을 수 있습니다.

결과적으로 동일한 사용자가 방문한 POI는 지리적으로 밀집되어 있는 경향이 있습니다. 우리는 사용자 체크인 활동에서 이러한 지리적 클러스터링 현상이 LBSN의 POI 추천에 활용될 수 있다고 믿습니다. 따라서 다음에서는 POI 추천에 활용하는 것을 목표로 POI에서 사용자 체크인 행동에 대한 이러한 지리적 영향을 연구하고 모델링합니다.

우리의 목표를 달성하기 위해 우리는 사용자 u_i 가 POI (j, k) 를 모두 체크인할 확률을 계산하겠습니다. 그림 2를 바탕으로 우리는 체크인 확률이 거듭제곱 법칙 분포를 따를 수 있다고 직관적으로 생각합니다. 그럼에도 불구하고, 동일한 사람이 먼 거리에 걸쳐 방문한 POI 쌍의 체크인 확률은 표준 거듭제곱 법칙 분포가 아니라는 것을 관찰했습니다. 그림의 왼쪽 부분이 선형적으로 감소(즉, 정규 규모에서 기하급수적으로 감소)하여 거듭제곱 분포에 매우 잘 들어맞음에도 불구하고, 오른쪽

부분은 때때로 불규칙하게 벗어날 수 있습니다(즉, 일부 지점에서 확률이 높습니다). 합리적인 설명은 사용자가 다른 장소로 여행하여 다중 체크인 공간 클러스터를 생성할 수 있다는 것입니다. 일반적으로 사용자의 체크인 POI가 가까운 거리에 있는 경향이 있다는 사실은 데이터 분석을 통해 확인되었습니다. 앞서 언급했듯이 인근 POI는 서로 더 관련성이 높으며 이는 강력한 지리적 영향력을 나타냅니다. 게다가 선형 부분은 그림 2의 플롯은 POI 쌍의 대부분(90%) 포함합니다. 따라서 우리는 동일한 사용자가 방문한 두 POI 사이의 거리에 대한 체크인 확률을 모델링하기 위해 거듭제곱 법칙 분포를 사용하여 다음과 같이 제안합니다.

$$y = a \times x^b \quad (5)$$

여기서 a 와 b 는 거듭제곱 분포의 매개변수이고, x 와 y 는 각각 동일한 사용자가 방문한 두 POI 사이의 거리와 체크인 확률을 나타냅니다. 방정식 (5)는 선형 모델에 맞게 'loglog' 척도로 방정식 (6)으로 변환할 수 있습니다.

$$\log y = w_0 + w_1 \log x \quad (6)$$

따라서 원래의 거듭제곱 법칙 분포는 다음 방정식을 통해 복구될 수 있습니다.

$$a = 2^{w_0} \quad b = w_1 \quad (7)$$

따라서 다음과 같이 선형 곡선 피팅 방법을 적용하면 회귀를 간단하게 구현할 수 있습니다. 보다 구체적으로, $y' = \log y$, $x' = \log x$ 라고 가정하면 다음과 같이 데이터를 피팅할 수 있습니다.

$$y'(x', \mathbf{w}) = w_0 + w_1 \cdot x' \quad (8)$$

여기서 w_0 과 w_1 은 선형 계수이며, 총칭하여 \mathbf{w} 로 표시합니다. 즉, 모델은 \mathbf{w} 의 형태로 학습할 수 있습니다. 과적합을 방지하기 위해 최소제곱 오차 방법으로 가중 계수에 접근하고 아래와 같이 계수가 큰 값에 도달하지 못하도록 페널티 항(즉, 정규화 항)을 추가합니다[6].

$$E(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^N \{y'(x'_n, \mathbf{w}) - t_n\}^2 + \frac{\lambda}{2} \|\mathbf{w}\|^2 \quad (9)$$

여기서 $E(\mathbf{w})$ 는 손실 함수를 나타내고, N 은 입력 데이터 세트의 카디널리티, t_n 은 x'_n 에 대응하는 기준 진실, λ 은 정규화 항입니다. 따라서 a 와 b 의 최적 값은 아래와 같이 손실 함수 $E(\mathbf{w})$ 를 최소화하는 설정을 형성합니다.

$$\text{opt}\{a, b\} = \arg \min_{a, b} E(\mathbf{w}) \quad (10)$$

다음에서는 LBSN에서 POI 추천을 구현하기 위해 Naive Bayesian 방식에 기반한 협업 추천 방법을 소개합니다. 주어진 사용자 u_i 와 그가 방문한 POI 집합 L_i 에 대해, L_i 의 모든 위치에서 체크인 활동을 한 확률을 다음과 같이 L_i 에 있는 POI의 쌍별 거리를 고려하여 정의합니다.

$$Pr[L_i] = \prod_{l_m, l_n \in L_i \wedge m \neq n} Pr[d(l_m, l_n)] \quad (11)$$

여기서 $d(l_m, l_n)$ 은 POI l_m 과 l_n 사이의 거리를 나타내며, $Pr[d(l_m, l_n)] = a \times d(l_m, l_n) \times b$ 는 위에서 얻은 포울로 분포 모델을 따릅니다. 여기서는 POI 쌍의 거리가 독립적이라고 가정합니다. 따라서 주어진 POI l_j (즉, 추천 후보), 사용자 u_i , 그리고 그녀가 방문한 POI 집합 L_i 에 대해 u_i 가 l_j 를 체크인할 확률은 다음과 같습니다.

$$\begin{aligned} Pr[l_j|L_i] &= \frac{Pr[l_j \cup L_i]}{Pr[L_i]} \\ &= \frac{Pr[L_i] \times \prod_{l_y \in L_i} Pr[d(l_j, l_y)]}{Pr[L_i]} \\ &= \prod_{l_y \in L_i} Pr[d(l_j, l_y)] \end{aligned} \quad (12)$$

POI 추천을 하려면 $L - L_i$ 의 모든 POI를 $Pr[l_j|L_i]$ ($l_j \in L - L_i$)에 따라 정렬하여 가장 높은 $Pr[l_j|L_i]$ 를 가진 POI를 사용자에게 반환합니다.

4. 통합된 협업 POI 추천

본 절에서는 아이디어 요소를 융합하는 협업 추천을 수행하기 위한 통합 프레임워크를 제안합니다. POI 추천에 있어서 사용자 선호도, 사회적 영향력, 지리적 영향력, POI의 등급을 예측하는 것과는 달리, 우리는 기존 정보 검색과 매우 유사한 후보 POI의 순위 목록을 반환하는 것을 목표로 합니다[4].

앞서 설명한 것처럼 사용자 선호도, 사회적 영향력, 지리적 영향력 등 각 요소를 활용하여 POI 추천을 구현할 수 있습니다. 따라서 직관적으로 세 가지 추천 시스템을 구현할 수 있습니다. 우리는 선형 융합 프레임워크를 사용하여 위에서 언급한 세 가지 추천자가 제공한 순위 목록을 최종 순위 목록으로 통합할 것을 제안합니다 [4, 21]. 여러 추천자를 통합함으로써, 각 추천 알고리즘에서 가장 높은 순위를 차지한 POI는 추천 시스템이 결과 목록의 상단에 사용자가 선호하는 POI를 고밀도로 배치하여 리콜(다른 높은 순위를 차지한 POI로 인해)과 정밀도를 모두 높일 수 있습니다.

$S_{i,j}$ 는 사용자 u_i 가 POI l_j 에서 체크인할 확률 점수를 나타내며, 즉, u_i 가 l_j 에서 체크인 활동을 할 가능성이 높을수록 $S_{i,j}$ 가 커집니다. 각각 사용자 선호도, 사회적 영향력, 지리적 영향력에 따른 추천 시스템에 해당하는 $S_{u,i,j}$, $S_{s,i,j}$, $S_{g,i,j}$ 는 POI l_j 에서의 사용자 u_i 의 체크인 확률 점수를 나타냅니다. $S_{i,j}$ 는 다음과 같습니다.

$$S_{i,j} = (1 - \alpha - \beta)S_{u,i,j}^u + \alpha S_{i,j}^s + \beta S_{i,j}^g \quad (13)$$

여기서 두 개의 가중치 파라미터 α 와 β ($0 \leq \alpha + \beta \leq 1$)는 사용자 선호도와 비교하여 사회적 영향력과 지리적 영향력의 상대적 중요성을 나타냅니다. 여기서 $\alpha = 1$ 은 $S_{i,j}$ 가 사회적 영향력에 기반한 예측에 전적으로 의존하고, $\beta = 1$ 은 $S_{i,j}$ 가 지리적 영향력에 기반한 예측에 전적으로 의존하며, $\alpha = \beta = 0$ 은 $S_{i,j}$ 가 사용자 선호도에만 의존한다는 의미입니다.

위의 융합 프레임워크에 따르면, 체크인 확률 점수 $S_{i,j}$ 를 추정하기 위해서는 사용자 선호도, 사회적 영향력, 지리적 영향력에 해당하는 $S_{u,i,j}$, $S_{s,i,j}$, $S_{g,i,j}$ 의 체크인 확률 점수를 각각 예측해야 합니다. 따라서 사용자 u_i 가 POI l_j 를 방문하기 위한 체크인 확률 $pu_{i,j}$, $ps_{i,j}$ 및 $pg_{i,j}$

ij를 추정하여 각각 Su_{ij} , Ss_{ij} 및 Sg_{ij} 를 구합니다. 먼저, 앞서 설명한 사용자 기반 협업 필터링의 아이디어를 기반으로 pu_{ij} 의 예측을 추정할 수 있습니다. 보다 구체적으로, 유사한 사용자들의 행동을 활용하여 방정식과 같이 예측을 실현합니다.

$$p_{i,j}^u = \frac{\sum_{u_k} w_{i,k} \cdot c_{k,j}}{\sum_{u_k} w_{i,k}} \quad (14)$$

여기서 $w_{i,k}$ 는 사용자 u_i 와 u_k 간의 유사성 가중치입니다. 마찬가지로, 친구 기반 협업 필터링의 아이디어를 기반으로 ps_{ij} 의 예측을 추정할 수 있습니다. 따라서 방정식 (3)에 따르면 다음과 같습니다.

$$p_{i,j}^s = \frac{\sum_{u_k \in F_i} SI_{k,i} \cdot c_{k,j}}{\sum_{u_k \in F_i} SI_{k,i}} \quad (15)$$

여기서 F_i 는 u_i 의 친구 집합이고, $SI_{k,i}$ 는 u_k 에서 u_i 까지의 사회적 영향력을 측정하는 가중치입니다.

마지막으로, pg_{ij} 는 방정식에서 직접 구할 수 있습니다.

$$p_{i,j}^g = Pr[l_j | L_i] = \prod_{l_y \in L_i} Pr[d(l_j, l_y)] \quad (16)$$

여기서 L_i 는 u_i 의 방문한 POI 집합이고, $d(l_j, l_y)$ 는 POI l_j 와 l_y 사이의 거리를 나타냅니다. 체크인 확률을 추정하고 나면 다음과 같이 해당 점수를 구할 수 있습니다.

$$\begin{aligned} S_{i,j}^u &= \frac{p_{i,j}^u}{Z_i^u}, \text{ where } Z_i^u = \max_{l_j \in L-L_i} \{p_{i,j}^u\} \\ S_{i,j}^s &= \frac{p_{i,j}^s}{Z_i^s}, \text{ where } Z_i^s = \max_{l_j \in L-L_i} \{p_{i,j}^s\} \\ S_{i,j}^g &= \frac{p_{i,j}^g}{Z_i^g}, \text{ where } Z_i^g = \max_{l_j \in L-L_i} \{p_{i,j}^g\} \end{aligned} \quad (17)$$

여기서 Z_{u_i} , Z_{s_i} , Z_{g_i} 는 정규화 용어입니다.

5. 경험적 평가

이 섹션에서는 제안된 협업 추천 알고리즘의 추천 품질을 협업 필터링 및 재시작을 통한 랜덤 워크를 포함한 일부 최신 추천 기술과 비교하고 몇 가지 흥미로운 질문을 조사하기 위해 여러 실험을 설계하고 수행합니다.

구체적으로 실험 설계는 다음과 같은 목표를 달성하는 것을 목표로 합니다.

- (1) 제안하는 방법은 사용자 선호도, 친구의 사회적 영향, 주변 위치의 지리적 영향을 고려하여 매개변수를 연구하고자 한다. α 그리고 β 최적의 권장 사항을 얻는 데 있어 위에 언급된 요소의 역할/가중치를 이해합니다.
- (2) 우리는 제안된 접근법의 효율성을 다른 최첨단 기술과 비교하여 우리의 아이디어를 검증하려고 합니다.
- (3) 친구의 사회적 영향에 대한 연구 관심이 높아지면서 두 친구 사이의 "사회적 유대"의 강도 측면에서 체크인 행동의 유사성을 더 연구할 계획입니다.
- (4) 우리 제안에서는 사용자 선호도를 발견하기 위해 사용자 기반 협업 필터링 접근 방식을 사용했습니다. 우리는 추천 품질을 더욱 향상시키기 위해 항목 기반 협업 필터링 접근 방식을 통합하는 타당성과 필요성을 탐색하려고 합니다.
- (5) 데이터 희소성이 LBSN의 POI 권장 사항에 어떤 영향을 미칠 수 있는지 이해하고 싶음.
- (6) 우리의 기술이 관심 분야 발견을 위한 체크인 기록이 많지 않은 콜드 스타트 사용자를 얼마나 잘 처리하는지

5.1 데이터세트 설명

가장 대표적인 LBSN인 Foursquare와 Whrrl의 웹사이트를 한 달간 크롤링하여 Foursquare에서는 사용자 153,577명, POI 96,229개, Whrrl에서는 사용자 5,892명, POI 53,432개로 구성된 두 개의 데이터세트를 수집했습니다. 우리의 성능 평가는 이 두 가지 대규모 실제 데이터 세트에 대해 수행됩니다. 체크인 기록을 요약한 후 사용자 POI 체크인 매트릭스 밀도를 4로 얻습니다.. 24×10^{-5} Foursquare 데이터세트 및 2.72×10^{-4} 각각 Whrrl 데이터 세트의 경우. 희소 데이터세트

(즉, 저밀도 사용자-POI 체크인 매트릭스)에 대한 추천 서비스의 효율성은 데이터세트에서 제공하는 정보가 제한되어 있기 때문에 일반적으로 높지 않습니다. 예를 들어 [12]에서 보고된 정밀도는 7로 사전 처리된 데이터 세트에 비해 0.17입니다. 8×10^{-4} 밀도. 따라서 실험에서는 알고리즘의 절대적인 효율성 측정 대신 상대적인 성능을 관찰하는 데 중점을 두었으며, 이는 LBSN 사용자 수가 계속 증가하고 더 많은 체크인 활동이 기록됨에 따라 개선될 것으로 예상됩니다. 평가를 용이하게 하기 위해 데이터 세트의 각 개별 사용자에게 대해 사용자가 방문한 모든 POI의 $x\%$ ($x = 10, 30, 50$ (기본값은 30))를 무작위로 표시합니다. 실험에서 평가된 POI 추천 알고리즘은 누락된 사용자-POI 쌍을 복구하는 데 사용됩니다.

5.2 성능 지표

평가 중인 POI 추천 알고리즘은 각 후보 POI(즉, 사용자가 방문하지 않은 POI)에 대한 순위 점수를 계산하고 상위 N 개의 최고 순위를 반환합니다. 타겟 사용자에게 추천하는 POI입니다. 예측 정확도를 평가하기 위해 반환된 POI 권장 사항에서 복구된 전처리 단계에서 이전에 표시한 POI 수를 알아내는 데 관심이 있습니다. 보다 구체적으로, (1) N 개의 추천 POI에 대한 복구된 POI의 비율, 그리고 (2) 전처리에서 삭제된 POI 집합에 대한 복구된 POI의 비율이라는 두 가지 지표를 조사합니다. 전자는 precision@N 반면 후자는 recall@N , 총칭하여 performance@N . 우리 실험에서는 다음과 같은 경우에 성능을 테스트합니다. $N=5, 10, 20$ 이고 기본값은 5입니다.

5.3 평가된 권장사항 접근 방식

사용자 선호도(U), 친구의 사회적 영향력(S), POI의 지리적 영향력(G)이라는 세 가지 요소가 통합된 협업 추천 알고리즘에 통합되며, 평가 시 USG로 표시됩니다. USG에서 $0 < \alpha, \beta < 1$ 의

가중치 매개변수를 제어하여 구성할 수 있는 여러 가지 최신의 새로운 협업 필터링 접근법도 비교를 위해 평가됩니다. USG 외에도 평가 중인 추천 접근 방식은 다음과 같습니다.

- **사용자 기반 CF(U로 표시)** - α 와 β 를 모두 0으로 설정한 USG의 특수한 경우입니다. 즉, 사용자 선호도만 추천에 고려됩니다.
- **친구 기반 CF(S로 표시)** - 이 역시 $\alpha = 1$ 로 설정하는 USG의 특수한 경우로, 타겟 사용자의 친구만 특정 추천에 사용됩니다. 앞서 소개한 것처럼 친구 간의 소셜 영향력 가중치를 도출하는 방법에는 두 가지가 있습니다. 하나는 식(4)[12]에 따라 친구를 기준으로 소셜 영향력 가중치를 계산하는 것이고, 다른 하나는 랜덤 워크 및 재시작 기법[24]을 사용하여 친구 간의 사회적 영향력 가중치를 도출하는 것입니다. 이 두 가지 접근 방식을 구분하기 위해 각각 S 와 S_{rwr} 로 표시합니다.
- **GI 기반 추천(G로 표시)** - 지리적 영향력 요인만 고려하는 이 접근 방식은 $\beta = 1$ 인 USG의 특수한 경우입니다. Random Walk with Restart- (RWR) - 최근 소셜 네트워크 기반의 협업 아이템 추천을 위해 개발된 최신 알고리즘입니다[12]. 랜덤 워크와 재시작을 통해 소셜 그래프와 사용자-아이템 매트릭스를 캡처한 그래프를 통해 아이템에 대한 사용자의 선호도를 예측합니다. - 사용자 선호도/사회적 영향력 기반 추천(US로 표시) - 사용자 선호도와 친구의 사회적 영향력을 모두 고려하는 이 방식은 $0 < \alpha < 1, \beta = 0$ 인 USG의 특수한 경우입니다.
- **사용자 선호도/지리적 영향력 기반 추천(UG로 표시)** - 사용자 선호도와 지리적 영향력을 모두 고려하는 이 방식은 $0 < \beta < 1, \alpha = 0$ 인 USG의 특수한 경우입니다.

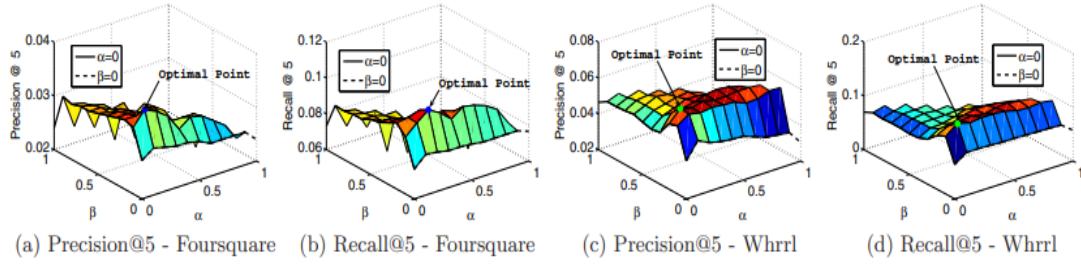


Figure 3: Tuning parameters

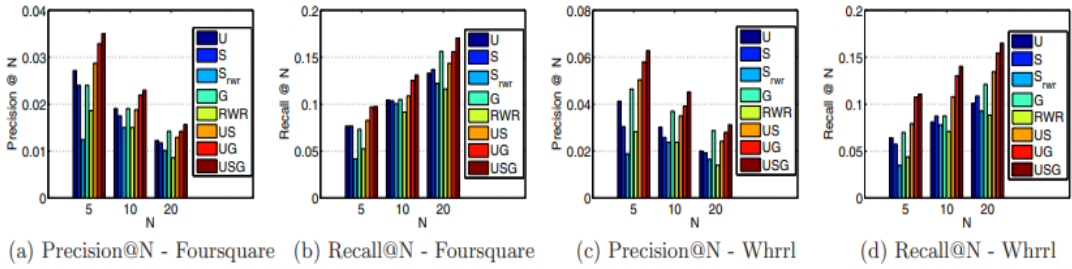


그림 4: 성능비교

5.4 튜닝 매개변수

앞서 언급했듯이, 두 개의 파라미터 α (사회적 영향 요인)와 β (지리적 영향 요인)를 제어하여 USG의 성능을 조정하고 평가를 위한 다른 추천 접근 방식으로 구성할 수 있습니다. 여기서는 사용자 선호도, 친구의 사회적 영향력, POI의 지리적 영향력이 최적의 USG 성능을 달성하는 데 어떤 역할을 하는지 이해하기 위해 USG에서 이를 다양하게 조정합니다. 마찬가지로 US에서는 α 를, UG에서는 β 를 조정하여 최적의 설정을 찾습니다. 그림 3은 다양한 α 및 β 설정에서 USG의 성능@5 결과를 보여 주며, 그림에 가장 적합한 파라미터 설정이 표시되어 있습니다. US와 UG의 최적 설정도 그림에서 확인할 수 있습니다. 즉, US는 점선, UG는 실선으로 표시되어 있습니다. 이러한 최적의 파라미터 설정은 표 1에도 요약되어 있습니다.

		Precision@5		Recall@5	
		α	β	α	β
Foursquare	US	0.1	—	0.1	—
	UG	—	0.2	—	0.2
	USG	0.1	0.1	0.2	0.1
Whrrl	US	0.1	—	0.1	—
	UG	—	0.1	—	0.1
	USG	0.1	0.2	0.1	0.1

표 1: 최적의 매개변수 설정

본 연구를 통해 우리는 최적의 추천에 사용자 선호도가 지배적인 역할을 하며, 사회적 영향력과 지리적 영향력은 모두 무시할 수 없다는 것을 쉽게 관찰할 수 있습니다. 보다 구체적으로, 표 1에서 볼 수 있듯이, 최고의 추천을 하는 데에는 사용자 선호도 요인이 최소 70% 이상 기여하고, 사회적 영향력과 지리적 영향력 모두 최소 10% 이상 기여합니다.

다음으로, 평가 중인 추천 접근법의 효과를 비교합니다. 그림 4는 모든 접근법의 최고 성능(즉, 최적의 매개변수 설정에서의 성능)을 $N(N = 5, 10, 20)$ 에서의 성능@N을 보여줍니다. 실험에는 Foursquare와 Whrrl 데이터 세트가 모두 사용되었습니다. 각각 그림 4(a)와 그림

4(b), 그림 4(c)와 그림 4(d)에 정밀도와 정확도가 표시되어 있습니다. 이 그림에서 USG는 모든 N 값에서 항상 정밀도와 정확도 측면에서 최고의 성능을 보이며 사용자 선호도, 사회적 영향력, 지리적 영향력의 세 가지 요소를 모두 결합했을 때의 강점을 보여줍니다. 실제 데이터 세트(즉, Foursquare와 Whrrl)는 모두 밀도가 낮다는 점에 주목하세요. 12]의 경험적 연구에 따르면, 사용자-항목 행렬의 밀도가 7.8×10^{-4} 인 사전 처리된 데이터 세트에 대해 보고된 정밀도는 약 0.17입니다. 따라서 사전 처리되지 않은 데이터 세트에서 측정된 낮은 정밀도는 합리적인 수준입니다. 가장 중요한 것은 USG가 두 데이터 세트 모두에서 기준 접근 방식 U(즉, 사용자 기반 CF)보다 약 50%의 성능 향상률을 보인다는 점입니다. 친구 기반 CF에 대한 두 가지 대안적인 사회적 영향력 측정 방법(즉, S와 Srwr)

사이에서 S가 Srwr보다 훨씬 더 나은 성능을 보인다는 것을 알 수 있습니다. 또한, 이 실험에서 RWR은 POI 추천의 경우 성능이 좋지 않은 것으로 나타났습니다. 이는 랜덤 워크 및 재시작 기법이 POI 추천에 적합한지에 대한 매우 흥미로운 의문을 제기합니다. 이후 섹션에서는 (i) 친구들 간의 체크인 행동의 유사성과 (ii) 친구들 간의 사회적 유대 사이의 상관관계를 분석하여 이 질문에 답하겠습니다. 나머지 실험에서는 친구의 사회적 영향력의 구성 요소로 US 및 USG에서 S를 사용합니다.

그림 4는 또한 사회적 영향력과 지리적 영향력을 모두 활용하여 POI 추천을 수행할 수 있음을 보여줍니다. 그림에서 볼 수 있듯이 S와 G 모두 U에 비해 비슷한 결과를 제공합니다.

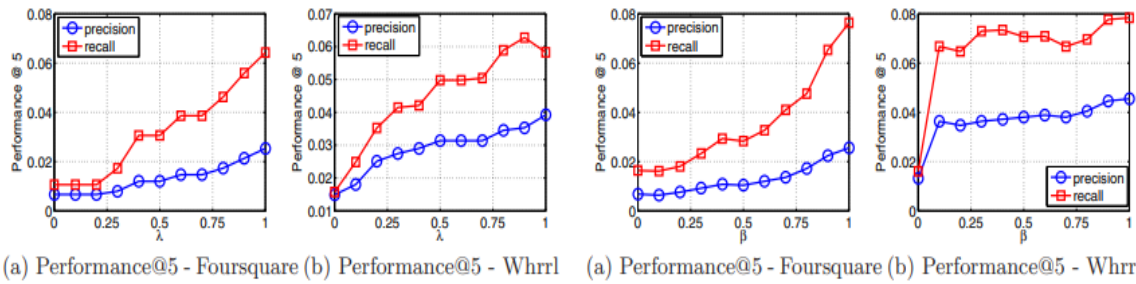


Figure 5: Fusion of U and L

Figure 6: Fusion of L and G

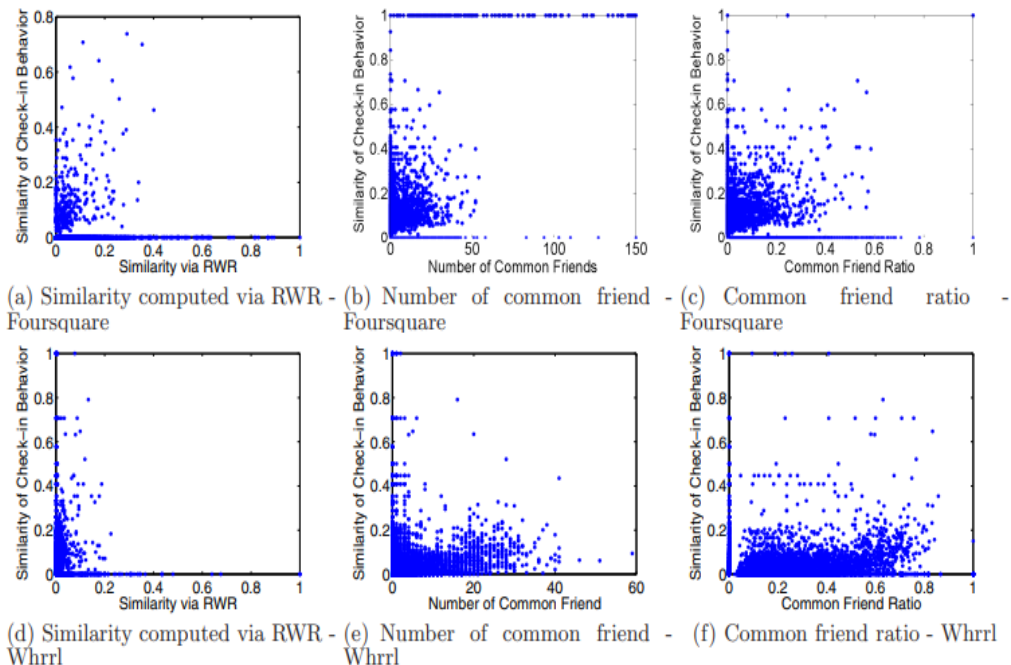


그림 7: 사회적 유대와 그것이 미치는 영향

LBSN에서는 체크인 활동이 사용자와 POI 간의 물리적 상호 작용을 포함하므로 지리적 영향이 매우 중요하며, 이는 연구에서 확인되었습니다. 위에서 살펴본 바와 같이 G는 일반적으로 S보다 성능이 뛰어나며 때로는 N이 20인 경우와 같이 U보다 더 나은 성능을 보이기도 합니다. 또한 UG는 항상 US보다 더 나은 성능을 보입니다. 이는 사용자 체크인 활동에서 나타나는 공간적 클러스터링 현상 때문입니다. 따라서 N이 상대적으로 클 경우 소셜 영향력에 기반한 사용자의 체크인 활동을 대부분 발견할 가능성이 매우 높습니다. Foursquare와 Whrrl 데이터 세트 모두에서 더 많은 요소를 고려할수록 성능이 더 좋은 것으로 나타났습니다. 예를 들어 US는 U 및 S보다, UG는 U 및 G보다, USG가 가장 좋은 성능을 보입니다.

5.6 아이템 기반 CF 연구

사용자 기반 CF 외에도 아이템 기반 CF는 사용자 대신 아이템 간의 유사도를 탐색하여 아이템에 대한 사용자의 선호도를 추정할 수 있습니다 [18]. 21]에서는 데이터 희소성 문제를 극복하기 위해 사용자 기반과 항목 기반 유사도를 모두 융합하는 CF 기법이 제안되었습니다[10]. 따라서 POI 추천을 위한 잠재적인 아이디어는 항목 기반 CF(L로 표시)를 사용하는 것입니다. 또한, POI 간의 영향력을 모델링하는 지리적 영향력은 아이템 기반 CF의 '항목 유사성'과 겹보기에는 유사해 보일 수 있습니다. 그러나 개념적으로 다르므로 혼동해서는 안 된다는 점을 알려드리고자 합니다. 이 섹션에서는 L을 UL 및 GL로 표시되는 새로운 접근 방식에 각각 U 및 G와 융합하면 U 및 G 단독보다 성능이 더 뛰어난지 검토하여 프레임워크에 L을 추가로 통합하는 아이디어를 살펴봅니다.

[21]과 유사하게, 울 가중치 매개변수 λ 을 도입합니다. $\lambda = 1$ 일 때, UL은 U로 축소되고; $\lambda = 0$ 일 때, L이 얻어집니다. 마찬가지로, GL에 가중치 매개변수 γ 를 도입합니다. 그림 5와 그림 6은 다양한 λ 및 γ 설정 하에서 Foursquare와 Whrrl 데이터셋에 대한 UL과 GL의 성능을 보여줍니다. 놀랍게도, 이 그림들은 POI 추천에서 L이 U나 L을 향상시키는 데 전혀 이점이 없음을 보여줍니다. 이는 우리의 응용 프로그램에서 아이템 기반 CF가 효과적인 접근법이 아니라는 것을 나타냅니다. 우리의 설명은 현재 단계에서 LBSN의 POI들이 아이템 기반 CF가 잘 작동하려면 충분히 많은 사용자에게 의해 방문되지 않았을 수 있다는 것입니다. 다시 말해, 두 POI 사이의 계산된 유사성은 사용자가 POI를 좋아할지 여부를 결정하는 좋은 실마리를 제공하지 못할 수 있습니다. U나 G만이 L보다 훨씬 더 나은 성능을 보이기 때문에, 우리는 우리의 추천 프레임워크에 L을 통합하지 않습니다.

5.7 사회적 영향에 관한 연구

앞서 살펴본 것처럼 Random Walk with Restart[12]는 POI 추천에 적합하지 않습니다. 그 이유를 종합적으로 파악하기 위해 사용자 체크인 행위의 유사도와 Random walk restart 기반으로 계산된 사용자 유사도 간의 상관관계를 분석합니다.

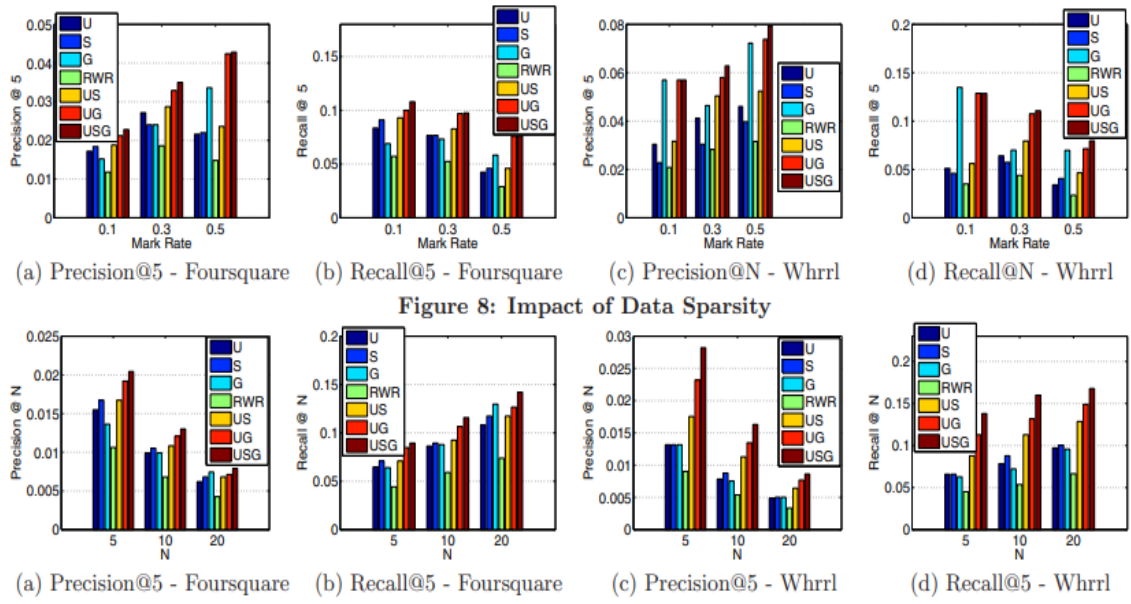


Figure 8: Impact of Data Sparsity

그림 9: 콜드 스타트 사용자를 위한 성능

[12]를 기반으로 소셜 그래프 매트릭스와 사용자 POI 체크인 매트릭스로부터 사용자 유사성을 도출할 수 있습니다. 그림 7(a)와 그림 7(d)는 최상의 RWR 설정에서 Foursquare 및 Whrrl 데이터 세트의 플롯을 보여줍니다. 두 그림 모두 유사한 사용자가 체크인 행동에서 반드시 높은 유사성을 갖는 것은 아니라는 것을 보여줍니다. 예를 들어, 유사도가 0.1 보다 큰 사용자 쌍은 일반적으로 Foursquare 및 Whrrl 데이터 세트 모두에서 체크인 동작에서 아무 것도 공유하지 않습니다. 결과는 사용자 친구의 취향이 실제로 크게 다를 수 있음을 나타내며 이는 최근 [16]에서도 논의되었습니다. 이 결과를 추가로 검증하기 위해 두 친구 간의 체크인 행동의 유사성과 사회적 유대 강도 간의 상관관계도 조사했습니다. 우리의 테스트에서 사회적 유대는 두 가지 형태로 정의됩니다: 1) 공통 친구 수(실험 결과는 그림 7(b) 및 그림 7(e) 참조) 및 공통 친구 비율(그림 7(c) 및 그림 7 참조) (f) 실험 결과), 공통 친구 비율은 Jaccard 계수로 측정됩니다. 매우 강

한 사회적 유대감을 갖고 있는 친구들(즉, 공통 친구 수가 많거나 공통 친구 비율이 더 높음)의 경우, 체크인 행동이 그림에 표시된 것처럼 반드시 유사하지는 않다는 것을 다시 한 번 알 수 있습니다. 위의 관찰을 통해 우리는 친구들의 취향이 다르다는 결론을 내렸습니다. 친구의 체크인 행동의 유사성이 반드시 사회적 유대의 강도 측면에서 반영되는 것은 아닙니다. 실제로 친구 간의 사회적 영향력을 측정할 때 사회적 유대 요인보다 체크인 행동 요인이 더 중요한 것으로 나타났습니다. 우리의 실험을 통해 S-S 알고리즘을 통해 최적의 설정을 발견했다. η 방정식 (4)에서 Foursquare 및 Whrrl 데이터 세트 모두에서 0.05 보다 작습니다. 이는 체크인 행동 요소가 사회적 유대 요소보다 더 중요하다는 것을 나타냅니다. RWR 두 요소를 동일하게 처리하는 동안, 마찬가지로 Foursquare 및 Whrrl 데이터 세트의 성능이 저하됩니다.

5.8 데이터 희소성의 영향

여기서는 방법을 연구합니다. USG 데이터 희소성 문제를 다룹니다. 서로 다른 희소성을 갖는 사용자-POI 체크인 행렬을 생성하기 위해 $x\% = 10\%, 30\%, 50\%$ 로 표시합니다. 그림 8과 같이 세 그룹의 테스트에 대한 원래 체크인 데이터 세트의 사용자 체크인 활동 기록. 마크오프 비율 x 가 클수록 즉, 사용자 POI 체크인 매트릭스가 더 희박합니다. 보여진 바와 같이, USG 모든 마크오프 비율에서 항상 최고의 성능@5을 나타냅니다. 특히 $x\% = 50\%$ 와 같이 데이터가 매우 희박한 경우 사용자에게 흥미로운 POI를 추천하는 데 지리적 영향이 매우 중요한 역할을 합니다. 그 이유는 사용자와 소셜 친구 모두 체크인 로그가 상대적으로 적기 때문입니다. 따라서 이러한 희박한 데이터에서 도출된 유사성 가중치나 사회적 영향 점수는 오해의 소지가 있을 수 있습니다. 반면, 지리에 의해 영향을 받는 글로벌 행동을 반영하는 지리적 영향은 LBSN에서 대부분의 사용자 행동에 적합합니다. 따라서 지리적 영향 요인을 통합하는 접근 방식은 다음과 같습니다. G, UG 그리고 USG, 다양한 데이터 희소성 시나리오에서 큰 강점을 보여줍니다.

5.9 콜드 스타트 사용자를 위한 테스트

마지막으로 콜드 스타트 사용자를 위한 POI 권장 사항의 성능을 테스트합니다. 여기서는 30% 체크인을 제거한 후 사용자-POI 체크인 매트릭스에서 체크인 활동이 5회 미만인 사용자를 콜드 스타트 사용자로 간주합니다. 그림 9에서 볼 수 있듯이 우리가 테스트한 모든 경우에 USG 항상 최고의 퍼포먼스를 보여줍니다. 콜드 스타트 사용자를 위한 POI 추천에서 사용자 선호도를 포착하기는 어려운데, 이는 해당 사용자가 방문한 POI가 몇 개 없기 때문입니다. 결과적으로, 사용자

선호도만을 고려하는 U는 최악의 성능을 보입니다. 사용자 체크인 활동의 공간적 클러스터를 탐색하는 G도 영향을 받습니다. 반면에, S는 사회적 친구들이 많은 유용한 체크인을 제공할 수 있기 때문에 사용자의 체크인 부족을 극복합니다. 이는 POI 추천에 잠재적으로 유용할 수 있습니다. 따라서, 이 실험에서 우리는 S의 추천 성능이 일반적으로 U와 G보다 더 잘 작동한다는 것을 발견합니다. 주목할 점은, 그림 8에서 극도로 희소한 시나리오에서 G의 성능이 S보다 나은 것을 발견했는데, 그 시나리오에서는 사회적 친구들의 체크인 기록도 매우 제한적이기 때문입니다. 따라서, 지리적 영향이 대부분의 사람들에게 적용 가능하기 때문에 우세합니다. 하지만, 콜드 스타트 사용자들에게 POI 추천을 위해 모든 세 요소가 매우 중요하다는 점은 주목할 만한데, USG가 항상 최고라는 것입니다.

6. 결론 및 향후 연구

본 연구에서는 위치 기반 소셜 네트워크에서 POI 추천 서비스를 활성화하고자 합니다. 우리의 아이디어는 사용자 선호도, 사회적 영향력, 지리적 영향력을 추천에 포함시키는 것입니다. 사용자 기반 협업 필터링을 통해 사용자 선호도를 도출하고 친구의 사회적 영향력을 포착하는 것 외에도 사용자 체크인 활동에서 공간적 클러스터링 현상을 밝히기 위해 거둬제곱 법칙 분포를 사용하여 POI 간의 지리적 영향을 모델링합니다. 또한, 우리는 사회적 영향력과 지리적 영향력을 지닌 POI에 사용자 선호도를 융합한 통합 POI 추천 프레임워크를 제안한다. Foursquare와 Whrrl에서 수집한 두 개의 대규모 실제 데이터 세트에 대해 포괄적인 성능 평가를 수행합니다. 실험 결과는 통합 협업 추천 기

술이 평가된 다른 모든 추천 접근 방식보다 우수하다는 것을 보여줍니다. 실험 결과 분석을 통해 추가 결과가 밝혀졌습니다.

1) 지리적 영향은 사회적 영향보다 POI 추천의 효과에 더 큰 영향을 미칩니다.

2) Random Walk and Restart 는 LBSN 의 POI 추천에 적합하지 않을 수 있습니다. 왜냐하면 친구는 상당히 다른 선호도를 보이기 때문입니다. (즉, 사회적 유대의 강도는 LBSN 의 사용자 간의 체크인 행동의 유사성을 반영하지 않습니다.

3) 아이템 기반 CF 는 현재 LBSN 상태에서 여러 위치에 대한 방문자 수가 부족하기 때문에 우리 애플리케이션에서 효과적인 접근 방식이 아닙니다.