

위치기반 추천 시스템의 벡터 분석에 의한 알고리즘 설계

배기성, 서송리, 석민수
성균관대학교 정보통신공학부
e-mail : aano_bae@hotmail.com

Design Algorithm of Location based Recommendation System by Vector Analysis

Keesung Bae, Songlee Suh, Minsoo Suk
School of Information and Communication Engineering,
Sungkyunkwan University

요 약

유비쿼터스 컴퓨팅 환경에서 추천시스템은 무수히 많은 정보들에 대하여 사람들이 적절한 선택을 할 수 있도록 도와준다. 사용자에게 필요한 정보를 찾아주고, 정보들의 우선순위를 결정해주는 추천시스템에 있어서 사용자의 위치는 보다 가치있는 정보를 제공할 수 있는 도구가 된다. 위치기반 추천시스템은 사용자가 아이템들로부터 얼마나 멀리 떨어져있는가를 고려하여 상위 리스트들을 제공할 수 있어야 한다. 하지만 일반적인 추천시스템에서 주로 사용되고 있는 기존의 사용자 기반 협업필터링 기법은 사용자의 자발적인 정보 입력에 의존함으로써 일정한 수의 사용자 정보가 축적되어 있지 않으면 정확한 추천이 불가능한 단점이 있다. 본 논문에서는 아이템에 기반한 협업 필터링 기법을 확률적으로 분석하고, 아이템의 위치에 따라 랭킹을 부여하는 방법과 사용자의 위치정보를 추천알고리즘에 적용시켜 보다 정확하고 효율적인 추천방법을 제안하였다.

1. 서론

정보통신과 멀티미디어 기술의 발전으로 인해 유비쿼터스 네트워크가 이미 우리 생활속에 깊이 들어와 있다. GPRS, 3G, Wi-Fi, Buletooth 등과 같은 무선 통신솔루션에 의한 서비스와 휴대전화, 휴대용 개인 단말기(PDA), 타블렛 PC, 노트북과 같은 이동형 단말기가 빠르게 보급되어 모바일 컴퓨팅 기술은 네트워크에 접속된 동안 이동성을 자유롭게 하였다. 그에 따른 사용자의 모빌리티 특성에 의한 서비스는 점점 확대되고 있다.

인터넷뿐만 아니라 우리 주변에서는 위치정보를 다룬 정보들과 실시간으로 결합하여 Map 서비스, 지리정보 서비스(GIS) 등 다양한 부가서비스들이 등장하게 되었는데 이것이 LBS(Location Based Service), 즉 위치기반 서비스이다. LBS는 이동통신망이나 GPS 등을 통해 개인이나 차량 등의 위치를 파악해 긴급구조, 교통정보 등을 서비스하는 신산업 분야로 향후 IT 산업

뿐만 아니라 전자상거래, 교통, 환경, 의료, 행정 등 사회 전 분야의 산업에 파급효과를 가져올 것으로 기대된다. 유비쿼터스 컴퓨팅 환경은 인간이 언제 어디서나 컴퓨팅 환경을 통해 편리한 생활을 영위하는데 목적이 있다. 따라서 위치에 따라 가장 개인화된 맞춤 서비스를 제공할 수 있다는 측면에서 LBS는 유비쿼터스 환경에서 더욱 중요한 서비스로 부각될 것으로 기대된다. 또 LBS는 유비쿼터스 컴퓨팅 환경에서 최고의 기능을 발휘할 수 있다. 왜냐하면 LBS를 실현하는데 가장 걸림돌이 되고 있는 정확한 위치정보의 획득 문제가 유비쿼터스 환경에서는 자연스럽게 해결될 수 있기 때문이다.

LBS는 보다 차별화되고 사용자의 취향에 맞는 서비스가 제공되도록 요구되어지고 있다. 이러한 요구를 수용하기 위해 인터넷에서는 다양한 서비스를 제공하지만 모바일 환경에서는 통신채널이나 단말기 하드웨어와 같은 여러 가지 제약사항으로 인해 각 이동통신

서비스 업체별로 제한적인 서비스만을 제공하고 있다.

현재 국내에서 제공하고 있는 위치기반 서비스는 주변 시설 정보 서비스, 교통 항법 서비스, 추적 서비스 등 현재의 위치정보를 이용한 서비스가 시행되고 있으며, 향후에는 최적 경로 선택 기능, 이동 차량의 동적 트래킹, 이동체 주변의 시설물 콘텐츠 검색, 부가 가치 시공간 분석, GPS 위치 데이터와 지도 데이터간의 맵매칭, 과거의 위치정보를 이용한 시공간 데이터 마이닝 등 고부가가치 정보 서비스 등으로 확대될 전망이다. 이러한 LBS 를 가능하게 하는 LBS 플랫폼에서는 이동체의 위치정보를 저장하고 관리하는 기능을 필수적으로 포함하고 있어야 하며, LBS 의 서비스의 질을 결정짓는 중요한 요소이다.

본 논문에서는 위치기반 추천시스템의 개념과 구성에 대하여 살펴보고 얼마나 많은 과거 정보들이 유용한가에 대해, 그리고 어떻게 가장 효과적으로 이용할 것인가에 대하여 아이템 기반 협업 필터링 알고리즘을 벡터 분석으로 사용자와 아이템의 위치 정보에 따른 추천 알고리즘을 제안하였다.

2. 위치기반 추천 서비스

2.1 LBS 기술의 정의

위치기반서비스(LBS, Location-Based Services) 기술은 유,무선 통신망 특히 이동 통신망 및 무선인터넷을 통해 모바일 환경에서 위치정보를 이용한 다양하고 부가적인 서비스를 제공하는 통신인프라를 활용한 서비스 기술로서 이 기술은 일반 대중을 위한 서비스뿐만 아니라 응급구조시스템, 재난관리시스템 등 공공 및 개인의 안전과 효율적인 자원의 관리를 위한 국가 정보 인프라로서 정보통신 기술 활용 등 사회적인 요구 증가에 부합되고 있는 차세대 서비스이다. 위치기반 서비스, 즉 LBS 는 기존의 유,무선 데이터 서비스 혹은 인터넷 서비스와 구별되는 새로운 개념의 서비스임에는 분명하지만 서비스의 사용자가 휴대형태의 단말을 사용하는 것이 보편화 되고 있는 상황에서 필연적으로 탄생되고 발전될 수 밖에 없는 서비스 영역이다. 따라서 LBS 를 일명 MLS, 즉 이동 위치정보 서비스라 부르기도 한다. 위에서 정의하는 바와 같이 위치기반 서비스는 아래 표와 같이 위치서비스와 이를 활용하는 응용서비스로 간단히 요약된다.

형태	Location Service (LCS)	Location-Based Service (LBS)
제공 서비스	휴대 단말의 지리적인 좌표 단위의 위치정보	휴대 단말의 위치 정보를 이용한 부가가치 서비스
서비스 제공자	이동통신망 사업자 이동통신망 임대 사업자 위치정보 제공 사업자	응용서비스 제공자 (위치기반 서비스 제공자 또는 위치 서비스 사용자 : 무선 서비스 제공자, 2차 ASP 사업자, 별정 이동통신 사업자, 이동통신망 임대 사업자)
서비스 이용자	위치기반 서비스 이용자	이동통신 사용자, 무선 서비스 제공자, 정부 기관, 군, 소방, 보안업체

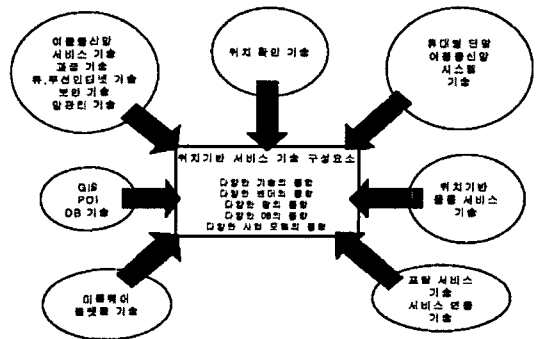
< 표 1. 위치서비스와 위치기반서비스의 관계 >

즉 LBS 는 아래와 같은 몇 가지 중요한 기술 요소들로 구성되며 이러한 요소들을 활용하여 사용자, 특히 이동형 단말을 휴대한 사용자에게 사용자의 위치나, 요구에 적합하고 최적화된 서비스를 간편한 방법에 의하여 실시간으로 제공함으로써 사용자의 위치정보 활용 요구를 만족시킬 수 있는 서비스를 의미한다.

- 사용자의 위치정보
- 단말 및 각종 데이터를 수집하는 장치의 위치 정보
- 이들 장치에서 입력되는 각종 음성, 영상, 감지 데이터
- 사용자, 단말, 데이터, 서비스 제공위치를 나타내는 지도
- 이들 정보를 가공하여 처리, 서비스를 제공하는 플랫폼

2.2 LBS 관련 기술과의 관계

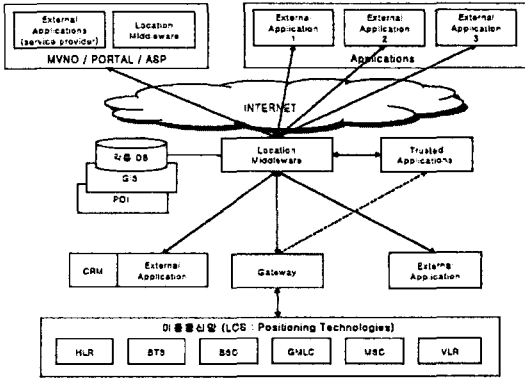
최근에 대두되고 있는 유비쿼터스 관련 기술 및 서비스는 이 위치기반 서비스와 밀접한 관계가 있으며 유비쿼터스를 구현하기 위한 기술은 통신 인프라 및 각종 데이터 수집을 위한 센서 기술등 현재 정보, 통신 관련 기술뿐만 아니라 정밀기계 관련된 기술까지를 총 망라한 넓은 영역을 포함하게 될 것이며 이러한 기반 기술을 접목하여 궁극적으로 사용자에게 제공하는 서비스는 기존의 정보통신 인프라를 이용하여 제공되던 서비스와 크게 다를 바가 없지만 가장 두드러지는 차이점은 위치기반 서비스에서 가장 중요한 요소인 위치정보가 기존의 서비스에 더 추가됨으로서 서비스의 활용도가 다양해지고 정보의 가치가 더 커지게 된다. 즉 유비쿼터스 환경에서의 서비스는 위치정보를 기반으로 하는 서비스 형태를 띠게 될 것이다.



< 그림 1. LBS 관련 기술 관계도 >

또한 위치기반 서비스는 거의 모든 경우에 이동형 단말을 이용하여 제공될 것이며 또한 이동망의 국제 로밍을 고려할 때 서비스의 글로벌라이징이 중요한 요소로 부각되고 있다. 특히 이동통신 인프라가 전 세계를 동작영역으로 하는 국제적인 로밍이 보편화 되는 경우에 위치기반 서비스는 그 활용도가 극대화될 것이며 전 세계를 다양한 이동통신 인프라에서 현재의 음성통화 서비스와 같이 위치기반 서비스도 상호연동되어 특별한 제약 없이 이용가능 하여야 한다. 이

런 요구 사항을 충족시키기 위하여 각종 정보의 표준화와 시스템 인터페이스의 표준화 정보접근 방법의 표준화 서비스 형태의 표준화, 특히 GIS 정보의 표준화가 무엇보다도 중요한 이슈로 등장하게 되었다.



< 그림 2. 위치기반 서비스 기술 관계도 >

위치기반 서비스를 제공하기 위하여 요구되는 기술들은 위의 그림에서 보인 바와 같이 이동 통신망을 이용한 서비스 제공에 필요한 모든 기술들이 망라되어야 한다. 여러 이해 당사자들이 관련되어 통합 제공되는 서비스 이므로 각 요소의 기술들이 연동될 수 있는 연동 기술이 중요한 의미를 가지게 되며 상호간의 표준화된 인터페이스를 정의하는 일이 무엇보다도 중요한 의미를 가지게 됨을 알 수 있다.

3. 협업 필터링 알고리즘

3.1 사용자 기반 협업 필터링 알고리즘의 문제점

과거에 사용자 기반 협업 필터링 알고리즘은 매우 성공적인 알고리즘이었다. 하지만 이 알고리즘의 광범위한 사용은 다음과 같은 문제점을 야기시켰다.

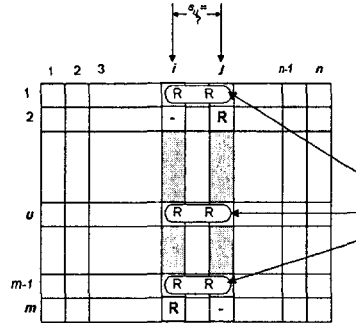
- Sparsity : 실제로 많은 광고 추천 시스템들이 거대한 양의 아이템 집합들을 평가하는데 이 알고리즘을 사용하여 왔다.(예를 들어 Amazon.com 의 도서 추천, CDnow.com 의 음악앨범 추천 등) 이들 시스템에서 적극적인 고객이 책을 구입할 확률은 아이템들중 1% 이하이다. 따라서 Nearest Neighbor 알고리즘을 이용하는 사용자 기반 추천시스템은 특별한 고객에게는 어떤 아이템 추천도 할 수 없을 것이다. 결과적으로 추천의 정확도는 매우 떨어지게 된다.
- Scalability : Nearest Neighbor 알고리즘은 사용자의 수와 아이템의 개수의 증가에 따라 많은 계산을 필요로 한다. 수백만의 사용자와 아이템을 보유한 전형적인 웹기반 추천 시스템들은 유지하는데 심각한 성능 저하를 유발하게 된다.

3.2 아이템 기반 협업 필터링 알고리즘

사용자 기반 협업 필터링과는 달리 아이템 기반 협업 필터링은 사용자가 선택한 아이템 집합이 무엇인가에 대하여 접근하고, 타겟 아이템 i 에 대해 사용자가 선택한 아이템과 얼마나 유사한가를 계산한다. 그리고 가장 유사한 k 개의 아이템 집합 $\{i_1, i_2, \dots, i_k\}$ 을 선택한다. 동시에 이 부합하는 유사도 집합 $\{s_{i1}, s_{i2}, \dots, s_{ik}\}$ 도 역시 계산된다. 일단 가장 유사한 아이템이 발견되면 그 때 이 유사하는 항목들 중 사용자가 선택한 아이템의 weighted 평균을 취하여 예측을 계산하게 된다. 이 두가지 측면은 아래에 유사도 계산과 예측 계산을 통해 설명된다.

3.2.1 유사도 계산

아이템 기반 협업 필터링 알고리즘에서 한가지 방법은 아이템들 사이의 유사도를 계산하고 그 때 가장 유사한 아이템을 선택하게 된다. 두 개의 항목들 i 와 j 사이의 유사도 계산의 기본적인 아이디어는 이들 아이템들을 모두 선택했던 사용자들을 처음에 따로 분리시켜놓고 유사도 s_{ij} 를 결정하기 위해서 유사도 계산 기법을 적용시키는 것이다.



< 그림 3. 상호 선택되었던 아이템들의 분리과 유사도계산 >

그림 3 의 아이템들 사이에서 유사도를 계산하는 방법은 다음과 같다.

- Cosine-based similarity

$$sim(i, j) = \cos(\vec{i}, \vec{j}) = \frac{\vec{i} \cdot \vec{j}}{\|\vec{i}\|_2 * \|\vec{j}\|_2}$$

< 식 1. ‘·’ 는 두 벡터들 간의 dot-product, $sim(i, j)$ 는 아이템 i 와 j 간의 유사도 >

이 경우, 2 개의 항목들은 m 차원 사용자 공간에서 2 개의 벡터들로 간주된다. 그들 사이의 유사성은 이 2 개의 벡터들 사이의 코사인 각을 계산하는 것에 의하여 측정된다.

- Correlation-based similarity

$$sim(i, j) = corr_{ij} = \frac{\sum_{u \in U} (R_{u,i} - \bar{R}_i)(R_{u,j} - \bar{R}_j)}{\sqrt{\sum_{u \in U} (R_{u,i} - \bar{R}_i)^2} \sqrt{\sum_{u \in U} (R_{u,j} - \bar{R}_j)^2}}$$

< 식 2. $R_{u,i}$ 는 아이템 i 에 대한 사용자 u 의 선택, R_i 는 i 번째 아이템의 선택의 평균, U 는 아이템 i 와 j 를 모두 선택한 사용자의 집합 > 이 경우, 2개의 항목 i, j 사이의 유사도는 Pearson-r correlation $corr_{i,j}$ 를 계산하는 것에 의해 측정된다.

• Adjusted Cosine similarity

$$sim(i, j) = \frac{\sum_{u \in U} (R_{u,i} - \bar{R}_u)(R_{u,j} - \bar{R}_u)}{\sqrt{\sum_{u \in U} (R_{u,i} - \bar{R}_u)^2} \sqrt{\sum_{u \in U} (R_{u,j} - \bar{R}_u)^2}}$$

< 식 3. \bar{R}_u 는 u 번째 사용자의 선택의 평균 >

3.2.2 예측 계산

협업 필터링 시스템에서 가장 중요한 단계는 예측에 관해서 출력 인터페이스를 생성하는 것이다. 우선 유사도 측정에 기반한 가장 유사한 아이템의 집합을 분리시켜 놓고, 그 다음 단계로 사용자가 선택한 아이템을 보고 예측들을 얻기 위해서 다음과 같은 기법을 사용하게 된다.

• Weighted Sum

$$P_{u,i} = \frac{\sum_{\text{all similar items, } N} (\theta_{i,N} * R_{u,N})}{\sum_{\text{all similar items, } N} (\theta_{i,N})} \quad < \text{식 4} >$$

이 기법은 그 이름과 같이 아이템 i 와 유사한 아이들들에 의해 주어지는 사용자의 선택을 더함으로써 사용자 u 에 대해 아이템 i 에 대한 예측을 계산한다. 근본적으로 이런 접근방식은 어떻게 적극적인 사용자가 유사한 아이들을 선택하는가를 저장한다.

• Regression

$$\bar{R}_i = \alpha R_u + \beta + \epsilon$$

<식 5 회귀모델 파라미터 α 와 β 는 선택벡터 모두를 조사하는 것에 의해 결정, ϵ 은 회귀 모델의 에러> 이 기법은 위의 weighted sum 기법과 유사한 접근이지만 회귀 모델을 기반으로한 선택의 근사화를 사용한다.

4. 위치기반 추천시스템 알고리즘

아이템에 대한 과거의 사용자의 선택이 데이터베이스에 축적되면 그 데이터는 사용자 선호 모델로 만들어질 수 있다. 또한 사용자 선호 모델은 아이들과 사용자 사이의 측정거리에 따라 표현되어지고, 과거의 사용자가 아이들을 선택한 것을 기반으로 위치에 의존한 아이들의 순위가 결정된다. 따라서 아이들 기반 협업 필터링 알고리즘과 위치정보를 결합하여 위치정보를 기반으로하는 추천시스템 알고리즘을 설계할 수 있다.

고유의 id를 가지는 s 개의 위치링크 $l=1, \dots, s$ 가 있다고 정의하고 서로 다른 사용자들에 의해서 N_l 번 방문된 위치 l 이 있다고 가정한다. 그리고 위치링크 l 이 선택되었던 물리적 위치를 나타내는 지점의 집합을 $P_l = (x_l, y_l)$, ($l \in N$)라고 하면, 위치링크 l 의 지역적 측정은 그 지점 P 를 중심좌표로 하는 타원을 계산함으로써 얻어진다.

위의 타원의 이차 방정식에서 이차 모멘트 행렬 (covariance matrix)인 M_l 을 이용하여 두 지점 $P=(x_p, y_p)$ 와 $Q=(x_q, y_q)$ 사이의 거리 $d(P, Q)$ 를 계산할 수 있다.

$$d_l(P, Q) = (x_p - x_q \quad y_p - y_q) M_l^{-1} \begin{pmatrix} x_p - x_q \\ y_p - y_q \end{pmatrix} \quad < \text{식 6} >$$

그리고 지점 P 에서 위치링크 l 에 대한 사용자의 선호도를 다음과 같이 정의하였다.

$$r_l(P) = \frac{1}{d_l(P, \bar{P}_l)} \quad < \text{식 7} >$$

여기서 거리함수 d_l 는 모든 위치가 다른 측정거리를 가지므로 위치의 수에 따라 색인되어진다.

5. 결론 및 향후 연구

본 논문에서는 Nearest Neighbor 계산 기법을 사용하는 사용자 기반 협업 필터링의 단점인 회소성 문제와 데이터베이스가 증가하면 성능이 저하되는 문제를 줄이기 위해 유사도 계산 기법과 예측 계산 기법을 사용하는 아이들 기반 협업 필터링을 사용하였다. 또한 모바일 환경에서 위치정보를 사용하는 위치기반 서비스에 적용할 수 있도록 아이들의 위치와 거리에 따른 사용자 선호도를 타원형 벡터 모델로 제시하여 위치기반 추천 시스템 알고리즘을 설계하였다. 타원형 거리 측정법을 사용한 알고리즘이 일반적인 Euclidian distance 기법과 Nearest Neighbor 기법보다 향상된 성능을 보여주었다. 하지만 아직도 개인의 프라이버시 문제와 악의 있는 사용자의 무분별한 아이들 선택, 시간성 고려 등과 같은 여러가지 해결해야 할 부분이 많이 남아있다. 향후에는 이런 문제점들을 보완하는 연구를 할 예정이다.

참고문헌

- [1] G. Karypis. Evaluation of item-based top-n recommendation algorithms. Technical report, University of Minnesota, Department of Computer Science / Army HPC Research Center, Minneapolis, USA, 2000.
- [2] Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., and Riedl, J. (1994). GroupLens: An Open Architecture for Collaborative Filtering of Netnews. In *Proceedings of CSCW '94*, Chapel Hill, NC.
- [3] Ling, C. X., and Li C. (1998). Data Mining for Direct Marketing: Problems and Solutions. In *Proceedings of the 4th International Conference on Knowledge Discovery and Data Mining*, pp. 73-79.
- [4] Aggarwal, C. C., Wolf, J. L., Wu K., and Yu, P. S. (1999). Horting Hatches an Egg: A New Graph-theoretic Approach to Collaborative Filtering. In *Proceedings of the ACM KDD'99 Conference*. San Diego, CA. pp. 201-212.
- [5] P. Bahl and V. N. Padmanabhan. RADAR: An in-building RF-based user location and tracking system. In *Proceedings of IEEE INFOCOM 2000*, pages 775-784, Mar. 2000.
- [6] S. Pradhan, C. Brignone, J.-H. Cui, A. McReynolds, and M. T. Smith. Websigns: Hyperlinking physical locations to the web. *IEEE Computer*, pages 42-48, Aug. 2001.