

# 유비쿼터스 환경에서 사용자 위치 기반의 개인화된 서비스 추천 알고리즘<sup>1</sup>

최정환, 장현수, 엄영익  
성균관대학교 정보통신공학부  
e-mail: {themars, jhs4071, yeiom} @ece.skku.ac.kr

## A Recommendation Algorithm for the Personalized Service Based on User Location in Ubiquitous Environments

Jung Hwan Choi, Hyun Su Jang, Young Ik Eom  
School of Information and Communication,  
Sungkyunkwan University

### 요 약

추천 서비스는 사용자에게 적합한 서비스를 선용적으로 제공하는 기술로써, 전자상거래 환경을 중심으로 널리 이용되고 있다. 그러나, 유비쿼터스 환경에서도 가장 활발한 기술 접목이 이루어지는 홈 네트워크 환경 내에 추천 서비스가 적용된 사례는 많지 않다. 본 논문에서는 홈 네트워크 환경에서 누적된 사용자와 기기 간 상호작용 정보들을 바탕으로 사용자 위치 기반의 개인화된 서비스를 추천하는 알고리즘을 제안한다. 본 알고리즘에서는 밀도기반 초기값 선정 기법을 적용한 군집화를 통해 필요한 데이터만을 추출함으로써 서비스 추천의 효율성 및 정확성을 높인다. 또한, 사용자 기반의 협업 필터링을 이용하여 데이터가 충분히 많지 않은 상황에서도 정확한 서비스 추천을 수행한다.

### 1. 서론

인터넷 쇼핑몰 등 전자상거래 환경에서는 이미 개인에게 적합한 상품들에 대한 추천 서비스가 널리 이용되고 있다. 고객의 필요와 욕구에 밀착된 서비스를 제공한다는 개인화의 취지에 비추어 볼 때, 이러한 추천 서비스는 점차 필수적인 요소로 자리매김 하고 있다[1]. 유비쿼터스 시대의 도래와 함께 개인화된 서비스에 대한 욕구가 날로 증대되고 있고[2], HP 사의 Cooltown[3], 카네기멜론 대학의 AURA[4] 프로젝트 등 사용자들의 욕구에 부응하는 여러 시스템들이 학교 및 기업체를 중심으로 연구되고 있다.

본 논문에서는 홈 네트워크 환경에서 누적된 사용자와 기기 간의 상호작용 정보들을 바탕으로 현재 위치에서 사용자에게 적합한 개인화된 서비스를 추천하는 알고리즘을 제안한다. 사용자와 기기 간의 상호작용 정보들은 프로파일 내에 누적되며, 이 정보들은 군집화 알고리즘을 통해 사용자의 위치에 해당하는 정보만 추출되어, 협업 필터링 기법을 적용하여 사용자에게 가장 적합한 서비스를 추천한다.

본 알고리즘은 모든 데이터를 분석하지 않고 군집화 기법을 이용하여 사용자가 위치한 군집 내의 데이터만을 분석함으로써 정확성 및 효율성을 증대시킨다.

또한, 사용자 기반의 협업 필터링을 이용함으로써 데이터가 충분히 많지 않은 상황에서도 보다 정확한 서비스 추천을 수행한다.

본 논문의 구성은 다음과 같다. 2 장에서는 배경 지식에 대해 알아보고, 3 장에서는 본 논문에서 제안하는 개인화된 서비스 추천 시스템 환경에 대해 설명한다. 4 장에서는 서비스 추천 알고리즘에 대해 알아보고 5 장에서 결론 및 향후 연구로 끝을 맺는다.

### 2. 관련 연구

#### 2.1. 상황 인식

상황 인식이란 어떤 사건이 발생하거나 어떤 행위를 하고자 할 때 주위의 다양한 환경 요소를 고려하여 다음에 취해야 할 행동이 능동적으로 결정되는 것을 말한다. 상황은 사용자의 현재 활동과 같이 개인적인 것일 수도 있으며, 현재 사용 중인 기기와 같이 기술적인 것일 수도 있으며, 온도, 위치, 시간과 같은 환경적인 것일 수도 있다[5].

상황인지 서비스를 위해 현재까지 가장 중점적으로 이용되는 상황은 사용자의 위치 정보이다. 올리베티 연구실의 Active Badge[6] 시스템을 시작으로 위치 인식을 위한 다양한 종류의 장치들이 개발되었으며, 이를 응용한 상황인지 서비스들이 현재까지도 꾸준히 연구되고 있다.

<sup>1</sup>“본 연구는 정보통신부 및 정보통신연구진흥원의 대학 IT 연구센터 지원사업의 연구결과로 수행되었음” (IITA-2007-(C1090-0701-0046))

2.2. 군집화 기법

군집화 기법은 통상 계층적 군집 기법과 파티셔닝 군집 기법으로 구분될 수 있다. 계층적 군집 기법은 가까운 개체들끼리 묶어감으로써 군집을 형성해 나가는 병합적 기법과 전체에서 2 개의 군집으로 나누는 것부터 시작하여 병합적 기법의 역순으로 진행되는 분할적 기법이 있다. 이러한 계층적 기법은 개체 수가 커지면 커질수록 계산에 매우 많은 시간과 용량을 필요로 하므로 적용에 제약이 따른다.

파티셔닝 군집 기법은 계층적으로 군집을 형성하지 않고 미리 규정된 개수만큼 분리시키는 기법이다. 대표적인 기법으로는 K-means 군집 기법이 있으며 방식은 (그림 1)과 같다. 계층적인 방법에 비해 부적절한 병합이 일어났을 때에는 재할당을 통해 극복할 수 있는 장점이 있지만, 초기값 선정에 따라 군집의 효율성 및 정확성이 결정된다는 단점이 있다[7].

<p><b>알고리즘: K-means 군집 기법</b>  <b>입력값:</b> 군집 개수 <math>k</math>, 데이터 <math>n</math>  <b>출력값:</b> 초기값 <math>k</math>  <b>방법:</b>                  (1) 군집 개수 <math>k</math> 만큼의 초기 값을 무작위로 선출한다.                  (2) 각 개체들의 거리를 계산하고 초기값과 가장 거리가 가까운 개체들을 군집으로 묶는다.                  (3) 군집의 중심값을 다시 계산하여 갱신한다.                  (4) 변화가 없을 때까지 (2)-(3)을 반복한다.</p>
---

(그림 1) K-means 군집 기법

2.3. 추천 기법

현재 대표적으로 많이 사용되고 있는 추천 기법에는 내용기반 추천과 협업 필터링 기반 추천이 있다. [8]. 내용기반 추천은 정보 검색분야에 바탕을 두고 있으며 사용자의 프로파일과 추천의 대상이 되는 아이템의 구성 요소를 비교하여 유사도가 높은 것들을 추천하는 기법이다. 협업 필터링은 크게 사용자 기반 협업 필터링과 아이템 기반 협업 필터링으로 나뉘어 지는데, 본 논문에서 적용되는 사용자 기반 협업 필터링은 추천의 대상이 되는 목표 사용자와 비슷한 프로파일을 가진 다른 사용자를 찾아 그 사용자가 선호하는 것은 본인 또한 선호할 것이라는 개념을 기반으로 추천하는 방식이다. 내용 기반의 추천 시스템은 시스템 초기에 높은 정확도를 나타내는 것으로 알려져 있지만, 데이터가 축적이 되면 협업 필터링의 정확도가 높아진다. 이러한 두 가지 방식의 장점들을 혼합한 혼합 추천 방식도 제안되었다[9].

3. 개인화된 서비스 추천 시스템 환경

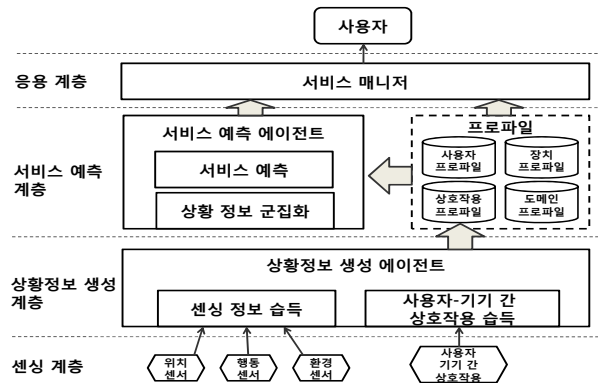
본 장에서는 홈 네트워크 환경에서 누적된 사용자와 기기 간의 상호 작용 정보들을 바탕으로 현재 위치에서 사용자에게 적합한 개인화된 서비스를 추천하는 시스템 환경에 대해 설명한다.

3.1. 시스템 구성도

(그림 2)는 개인화된 서비스 추천 시스템의 전체 구성도이다. 센싱 계층, 상황정보 생성 계층, 서비스 예측 계층, 응용 계층의 4 계층으로 이루어진 시스템은 계층 별 순차적 처리를 통해 사용자에게 개인화된 서

비스를 추천한다.

센싱 계층에서는 홈 네트워크 환경 내에 구축된 위치, 행동, 그리고 환경 센서 등으로부터 각종 상황 정보들을 취득한다. 또한, 사용자와 기기 간 상호작용도 파악한다. 센싱된 정보들은 상황정보 생성 계층의 상황정보 생성 에이전트를 통해 하나의 상호작용 상황 정보로 취합되며, 이는 상호작용 프로파일 내에 누적된다. 서비스 예측 에이전트에서는 상호작용 프로파일 및 기타 다른 프로파일들의 정보를 토대로 군집화 및 예측 과정을 통해, 사용자의 현재 위치에 가장 적합한 서비스를 예측한다. 이 정보는 최종적으로 서비스 매니저를 통해 사용자에게 추천된다.



(그림 2) 개인화된 서비스 추천 시스템의 구성

3.2. 상황정보 생성 에이전트

상황정보 생성 에이전트에서는 각종 센서로부터 얻은 센싱 정보들을 취합하여 하나의 상호작용 상황 (Interaction Context, IC)을 생성하고, 이를 상호작용 프로파일 내에 저장한다.



(그림 3) 상호작용 상황(IC)의 형식

(그림 3)는 IC의 데이터 형식을 나타낸다. 각종 센서들로부터 받은 센싱 정보의 요소  $s_1...s_n$ 는 사용자와 기기 간 상호작용 정보 요소  $i_1...i_n$ 와 취합해 하나의 IC를 생성한다. 이렇게 생성된 IC들은 사용자의 행동 패턴을 나타내며, 누적된 정보들은 서비스 예측의 지표로 활용된다.

3.3. 서비스 예측 에이전트

서비스 예측 에이전트는 상황정보 생성 에이전트를 통해 생성된 IC들을 분석하여 사용자에게 적합한 서비스를 예측한다. 우선 군집화 과정을 통해 누적된 수많은 IC들에 대한 군집을 형성한 후, 현재 사용자가 위치한 곳에 해당하는 군집 내의 정보를 기반으로 서비스 예측을 수행한다. 예측된 서비스는 최종적으로 서비스 매니저를 통해 사용자에게 추천된다. 군집화 및 서비스 예측에 관한 기술적인 알고리즘은 4장에서 설명한다

4. 서비스 예측을 위한 알고리즘

본 장에서는 서비스 예측 에이전트 내에서 수행하는 상황정보 군집화 및 서비스 예측을 위한 알고리즘에 대해 설명하고, 실 데이터를 적용하여 본 알고리즘의 성능을 분석한다.

4.1. 상황정보 군집화 알고리즘

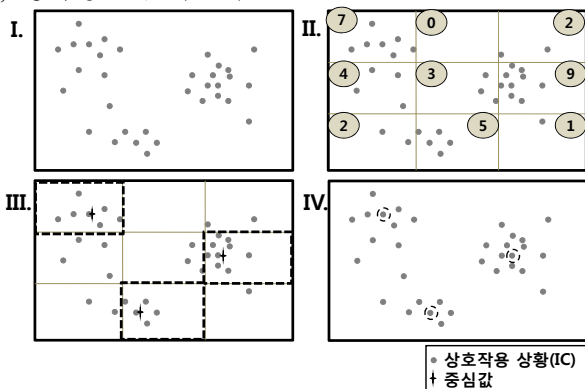
상황정보 생성 에이전트로부터 생성된 IC 들은 상호작용 프로파일 내에 누적되어 예측 서비스에 활용된다. 그러나, 서비스 예측을 위해 수 많은 IC 들을 모두 분석하기에는 효율성이나 정확성 측면에서 좋지 않다. 따라서, 군집화 알고리즘을 통해 현재 사용자의 위치에 근접한 군집 내의 데이터만 분석함으로써 예측의 효율성 및 정확성을 높인다.

군집화 기법으로 2.2 절에 언급한 K-means 군집화 기법을 이용한다. 이 기법의 전통적인 문제점으로 지적되었던 초기값 선정 문제를 해결하기 위해 밀도 기반 초기값 선정 방식을 적용하여 그 효율성을 높인다.

<p><b>알고리즘:</b> 밀도 기반 초기값 선정 기법</p> <p><b>입력 값:</b> 군집 개수 <math>k</math>, 데이터 <math>n</math>, 도메인 상황정보</p> <p><b>출력 값:</b> 초기값 <math>k</math></p> <p><b>방법:</b></p> <ol style="list-style-type: none"> <li>(1) 도메인을 <math>k^*k</math> 수로 균등 분할한다.</li> <li>(2) 분할된 파티션 내의 밀도를 계산한다.</li> <li>(3) 군집 개수 <math>k</math> 만큼 밀도의 크기가 큰 파티션을 오름 차순으로 정렬한다.</li> <li>(4) 해당 <math>k</math>개의 파티션에서 각 데이터들의 중심 값을 구한다.</li> <li>(5) 각각의 중심 값을 기준으로 가장 가까운 값을 초기값 <math>k</math>로 선정한다.</li> </ol>
--

(그림 4) 밀도 기반 초기값 선정 알고리즘

(그림 4)는 밀도 기반 초기값 선정 알고리즘을 나타낸다. 기존의 초기값 선정 방식은 단순히 임의의  $k$  개를 선정하는 방식이므로, 선정된 초기 값이 한쪽에 치우치게 되면 군집화를 이루는 과정에 보다 많은 비용을 소모하게 되고, 결과도 정확하지 않을 수 있다. 하지만, 밀도 기반 초기값 선정 알고리즘을 적용하면, 데이터가 많이 분포된 곳을 중심으로 초기값 선정이 이루어지므로, 군집화를 이루기 위한 비용은 적어지고, 정확성은 높아진다.



(그림 5) 밀도 기반 초기값 선정 과정

(그림 5)는 군집 개수  $k$  가 3 일 때, 위의 알고리즘에 맞추어 초기값이 선정되는 과정을 보인 것이다. 각 파티션의 밀도를 계산하여, 가장 밀도가 높은 파티션을  $k$  개 선정하여 그 안의 중심 값에 가장 가까운 값이 초기값으로 선정된다.

4.2. 협업 필터링을 이용한 예측 알고리즘

상황정보 군집화를 통해 군집이 이루어지면, 현재 사용자의 위치에서 가장 근접한 군집 내의 IC 를 대상으로 서비스 예측을 수행한다. 사용자가 위치한 군집 내의 IC 의 수가 많고 적음에 따라 두 가지 방향으로 예측이 이루어진다. 여기서 데이터의 많고 적음은 미리 정해진 한계값(T)에 의해 판별한다.

Case 1. 군집 내 IC 가 충분히 많은 경우 (IC>=T)

$$P_{U,j} = \frac{I_{U,device(j)} * ratio_{device(j)}}{\sum_{IC \in C} I_{U,device(j)}} \dots\dots\dots (1)$$

식 (1)에서  $P_{U,j}$ 는 사용자  $U$ 의 기기  $j$ 에 대한 선호도를 나타낸다. 선호도는 사용자  $U$ 가 기기  $j$ 와 상호작용을 한 횟수에 기기  $j$ 의 ratio 값을 곱한 것을 군집  $C$  내에서 이루어진 모든 IC 로 나눈 값으로 결정된다. 여기서 기기  $j$ 에 대한 ratio 값은 기기들의 동작 수에 따라 각각 상이하게 책정이 되는 값으로 식 (2)와 같이 동작 수에 대한 비율로 정해지며, 기기 프로파일 내의 각각의 기기 정보 내에 정의된다.

$$ratio_{device(j)} = \frac{1}{Number\ of\ Actions_{device}} \dots\dots\dots (2)$$

이렇게 결정된 사용자와 기기 간의 선호도 중 가장 높은 선호도를 가진 기기들을 우선 순위로 정렬하여 사용자에게 추천한다.

Case 2. 군집 내 IC 가 충분하지 않은 경우(IC<T)

IC 가 충분하지 않은 상태에서 사용자에게 정확한 예측 서비스를 제공하기란 쉽지 않다. 따라서, 이 경우에는 사용자 자신과 가장 유사한 사용자를 찾아 그의 데이터를 서비스 추천에 활용하는 사용자 기반의 협업 필터링이 효과적이다.

우선, 사용자  $U$ 와 가장 유사한 사용자  $J$ 를 찾기 위해 이용성향 요소( $f_1..f_n$ ) 값을 기준으로 식 (3)의 피어슨 상관관계수[10]로 유사성 있는 사용자를 찾는다.  $U_i$ 는 사용자  $U$ 의 이용성향 요소  $f_i$ 이며,  $\bar{U}$ 는 사용자  $U$ 의 이용성향 요소들의 평균값이다.  $J$ 의 경우도 동일하다. 결과값은 유사할수록 1에 가까우며, 유사하지 않을수록 0, 배타적인 값일 경우 -1에 가까운 값이 나타난다. 여기서 배타적인 값은 고려하지 않으므로, 절대값을 취해 항상 양의 값이 나오도록 한다.

$$r_{UJ} = \frac{\sum_{IC \in C} (U_i - \bar{U})(J_i - \bar{J})}{\sqrt{\sum_{IC \in C} (U_i - \bar{U})^2} * \sqrt{\sum_{IC \in C} (J_i - \bar{J})^2}} \dots\dots\dots (3)$$

모든 사용자에게 식 (3)을 적용하여 유사도를 측정하고, 이 중 가장 큰 유사도를 가진 사용자는 가장 유사한 사용자  $J$ 가 되며,  $J$ 의 정보를 참고로 하여 사용자  $U$ 에 대한 선호도가 계산된다.

$$P_{U,j} = \frac{I_{U,device(j)} * ratio_{device(j)}}{\sum_{IC \in C} I_{U,device(j)} + |r_{UJ}| * \frac{I_{J,device(j)} * ratio_{device(j)}}{\sum_{IC \in C} I_{J,device(j)}} \dots\dots\dots (4)$$

식 (4)와 같이 유사한 사용자  $J$ 의 선호도를 참고함으로써 IC 가 충분하게 많지 않은 경우라도 사용자에게 보다 정확한 서비스 추천이 가능하다.

4.3. 실 데이터를 통한 알고리즘 성능 분석

본 절에서는 앞서 설명한 알고리즘에 실제 수치를 대입하여 계산한 결과를 제시하고 성능을 분석한다.

<표 1> 군집 내 사용자와 기기 간 상호작용 예시

기기(d) \ 사용자(U)	$U_1$	$U_2$	$U_3$
TV ( $d_1$ )	24	127	42
Radio ( $d_2$ )	21	84	62
MP3 ( $d_3$ )	19	42	50
총계	64	253	154

현재 위치에서 사용자  $U_1$ 에게 가장 적합한 서비스 추천을 수행하려고 한다면, 우선, 군집화를 수행하고, 사용자  $U_1$ 의 위치와 가장 근접한 군집을 선택하여 그 안의 IC 들을 대상으로 예측을 수행한다. IC가 충분한지 아닌지를 결정하는 한계값(T)이 70이라고 할 때, <표 1>에서  $U_1$ 의 데이터는 충분하지 못하므로, 가장 유사한 사용자의 데이터를 참조하는 사용자 기반의 협업 필터링을 이용하여 서비스 추천을 수행한다.

<표 2>는 사용자의 이용성향을 분석한 이용성향 요소 ( $f_1...f_n$ )에 대한 값이다. 각각의 이용성향 요소는 사용자의 전체 상호작용 중 특정 기기를 상호작용한 비율로 구해진다. 식 (3)의 공식을 이용해  $U_1$ 을 기준으로 하여,  $U_2, U_3$  각각의 유사도를 측정한다.

<표 2> 사용자 프로파일 내의 성격 요소 예시

성격요소(f) \ 사용자(U)	$U_1$	$U_2$	$U_3$
$f_1$	0.20	0.20	0.12
$f_2$	0.16	0.12	0.08
$f_3$	0.16	0.16	0.04
$f_4$	0.20	0.16	0.04

<표 3> 유사도 측정 결과 값

유사도(r)	값
$r_{U_1,U_2}$	0.707
$r_{U_1,U_3}$	0.302

계산 결과 <표 3>에서와 같이 사용자  $U_1$ 과  $U_2$ 가 보다 유사성이 있음을 입증할 수 있다.  $U_2$ 가 행한 상호작용 상황 정보는  $U_1$ 의 데이터가 충분치 않으므로  $U_2$ 의 데이터를 활용하여 추천한다. 기기 프로파일 내에 정의된 ratio 값과 앞서 판별한 유사한 사용자의 정보를 바탕으로 식 (4)를 이용해 계산한다.

<표 4> 협업 필터링을 적용한 기기별 사용자 선호도

기기	본인선호도고려	유사사용자참조	Ratio
TV( $d_1$ )	0.01875	0.03649	1/20
Radio( $d_2$ )	0.01563	0.02681	1/21
MP3( $d_3$ )	0.01856	0.02589	1/16

<표 4>에서와 같이, 본인의 데이터만을 사용하는 것보다는 유사 사용자를 참조하는 것이 선호도 간의 구분이 크므로 보다 명확하고 신뢰할 수 있는 결과를

얻을 수 있다. 최종적으로, 선호도 값의 우선 순위에 따라 TV, Radio, MP3 순으로 서비스가 추천된다.

5. 결론

본 논문에서는 홈 네트워크 환경 내에서 사용자와 기기 간의 상호작용 정보들을 바탕으로 현재 위치에서 가장 적합한 서비스를 추천하는 알고리즘을 제안하였다. 대량의 정보들을 모두 분석하는 대신에 군집화 기법을 통해 현재 사용자의 위치에 근접한 군집 내의 데이터만을 이용함으로써 예측의 효율성 및 정확성을 향상시켰다. 또한, 사용자 기반 협업 필터링 기법을 이용함으로써 데이터가 충분하지 않은 상황에서도 보다 정확한 서비스의 추천이 가능하였다.

본 알고리즘은 홈 네트워크 환경뿐만 아니라, 기업, 학교 등 다양한 유비쿼터스 환경에 적용될 것으로 기대된다. 향후 연구로는 시간 정보 등과 같은 보다 다각적인 상황정보들을 활용하여 한층 정확한 서비스를 추천하는 개선된 방법에 대한 연구가 진행되어야 할 것이며, 실제로 본 알고리즘을 적용한 컴포넌트를 에이전트 기반 미들웨어 내에 적용하는 연구가 진행되어야 할 것이다.

참고문헌

- [1] W. Lee, "Towards agent-based decision making in the electronic market place: Interactive recommendation and automated negotiation," Elsevier Expert Systems with Applications Vol. 27, pp. 665-679, 2004
- [2] O. Kwon, K. Yoo, and E. Suh, "UbiDSS: a proactive intelligent decision support system as an expert system deploying ubiquitous computing technologies," Elsevier Expert Systems with Applications Vol. 28, pp.149-161, 2005
- [3] J. J. Barton and T. Kindberg "The Cooltown User Experience," Proceedings of CHI2001, pp. 149-156, 2001
- [4] D. Garlan, D. P. Siewiorek, A. Smailagic, and P. Steenkiste "Project Aura: Toward Distraction-Free Pervasive Computing," IEEE Pervasive Computing, 2002
- [5] S. Cuddy, M. Katchabaw, and H. Lutfiyya, "Context-aware Service Selection based on Dynamic and Static Service Attributes," WiMob'2005, IEEE International Conference, pp.13-20, 2005
- [6] R. Want, A. Hopper, V. Falcao, and J. Gibbons, "The Active Badge Location System," ACM Transactions on Information Systems, pp. 91-102, 1992
- [7] Q. He, "A Review of Clustering Algorithms as Applied in IR," UIUCLIS, 1999
- [8] M. Balabnovic and Y. Shoham, "Content-based, Collaborative Recommendation," Communications of the ACM Vol. 40, pp. 66-72, 1997
- [9] M. Claypool, A. Gokhale, T. Miranda, P. Murnikov, D. Netes and M. Sartin, "Combining content based and collaborative filters in an online newspaper," Proceedings of ACM SIGIR Workshop on Recommender Systems, 1999
- [10] D. Billsus, and M. J. Pazzani, "Learning Collaborative Information Filters," Proceedings of ICML, pp. 46-53, 1998